



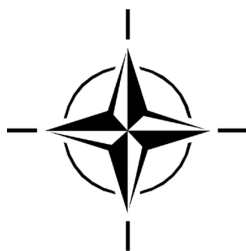
STO TECHNICAL REPORT

TR-HFM-290

Advances in Military Personnel Selection

(Progrès dans la sélection du personnel militaire)

Final report of Task Group HFM-290.



Published May 2022



NORTH ATLANTIC TREATY
ORGANIZATION



AC/323(HFM-290)TP/1076

SCIENCE AND TECHNOLOGY
ORGANIZATION



www.sto.nato.int

STO TECHNICAL REPORT

TR-HFM-290

Advances in Military Personnel Selection

(Progrès dans la sélection du personnel militaire)

Final report of Task Group HFM-290.

The NATO Science and Technology Organization

Science & Technology (S&T) in the NATO context is defined as the selective and rigorous generation and application of state-of-the-art, validated knowledge for defence and security purposes. S&T activities embrace scientific research, technology development, transition, application and field-testing, experimentation and a range of related scientific activities that include systems engineering, operational research and analysis, synthesis, integration and validation of knowledge derived through the scientific method.

In NATO, S&T is addressed using different business models, namely a collaborative business model where NATO provides a forum where NATO Nations and partner Nations elect to use their national resources to define, conduct and promote cooperative research and information exchange, and secondly an in-house delivery business model where S&T activities are conducted in a NATO dedicated executive body, having its own personnel, capabilities and infrastructure.

The mission of the NATO Science & Technology Organization (STO) is to help position the Nations' and NATO's S&T investments as a strategic enabler of the knowledge and technology advantage for the defence and security posture of NATO Nations and partner Nations, by conducting and promoting S&T activities that augment and leverage the capabilities and programmes of the Alliance, of the NATO Nations and the partner Nations, in support of NATO's objectives, and contributing to NATO's ability to enable and influence security and defence related capability development and threat mitigation in NATO Nations and partner Nations, in accordance with NATO policies.

The total spectrum of this collaborative effort is addressed by six Technical Panels who manage a wide range of scientific research activities, a Group specialising in modelling and simulation, plus a Committee dedicated to supporting the information management needs of the organization.

- AVT Applied Vehicle Technology Panel
- HFM Human Factors and Medicine Panel
- IST Information Systems Technology Panel
- NMSG NATO Modelling and Simulation Group
- SAS System Analysis and Studies Panel
- SCI Systems Concepts and Integration Panel
- SET Sensors and Electronics Technology Panel

These Panels and Group are the power-house of the collaborative model and are made up of national representatives as well as recognised world-class scientists, engineers and information specialists. In addition to providing critical technical oversight, they also provide a communication link to military users and other NATO bodies.

The scientific and technological work is carried out by Technical Teams, created under one or more of these eight bodies, for specific research activities which have a defined duration. These research activities can take a variety of forms, including Task Groups, Workshops, Symposia, Specialists' Meetings, Lecture Series and Technical Courses.

The content of this publication has been reproduced directly from material supplied by STO or the authors.

Published May 2022

Copyright © STO/NATO 2022
All Rights Reserved

ISBN 978-92-837-2394-3

Single copies of this publication or of a part of it may be made for individual use only by those organisations or individuals in NATO Nations defined by the limitation notice printed on the front cover. The approval of the STO Information Management Systems Branch is required for more than one copy to be made or an extract included in another publication. Requests to do so should be sent to the address on the back cover.

Table of Contents

	Page
List of Figures	ix
List of Tables	x
Acknowledgements	xi
HFM-290 Membership List	xii
Executive Summary and Synthèse	ES-1
Chapter 1 – An Introduction to Advances in Military Personnel Selection	1-1
1.1 Background	1-1
1.2 Relevance of Topics	1-2
1.2.1 Integrity-Related Assessments	1-2
1.2.2 Online/Computerised Testing	1-3
1.2.3 Diversity	1-3
1.3 Closing Remarks	1-4
1.4 References	1-4
Chapter 2 – History of Integrity Testing	2-1
2.1 Introduction	2-1
2.2 Personality Testing (1914 – 1945)	2-1
2.3 Cold War Research (1945 – 1970s)	2-4
2.4 Changing Direction (1970s – 1990s)	2-7
2.5 Concluding Remarks	2-9
2.6 References	2-9
Chapter 3 – Current State of Military Practice and Research on Integrity Assessment	3-1
3.1 Introduction	3-1
3.2 Current Practice	3-1
3.2.1 Methods of Integrity Assessment	3-2
3.2.2 Contribution to the Selection Decision	3-2
3.2.3 Difficulties and Challenges with Integrity Assessment	3-4
3.3 Ongoing Integrity Research	3-4
3.3.1 Belgium	3-5
3.3.2 Canada	3-5
3.3.3 Finland	3-6
3.3.4 Sweden	3-7

3.4	Conclusion	3-8
3.5	References	3-9
Chapter 4 – Measuring Integrity		4-1
4.1	Existing Approaches to Assess Integrity	4-1
4.1.1	Overt and Covert Measures	4-1
4.1.2	Other Measures	4-1
4.1.2.1	Polygraph Testing	4-1
4.1.2.2	MMPI	4-2
4.2	Commercial Integrity Tests and Measures	4-2
4.2.1	Summary	4-15
4.3	Emerging Approaches to Assessing Integrity	4-15
4.3.1	Platform-Centric Approaches	4-15
4.3.2	Information-Centric Approaches	4-16
4.3.3	Implications	4-17
4.3.3.1	Theoretical Relevance	4-17
4.3.3.2	Applicant Reactions	4-18
4.3.3.3	Job-Relevance	4-18
4.3.3.4	Lack of Standardisation	4-18
4.3.3.5	Legal Issues	4-19
4.4	Concluding Remarks	4-19
4.5	References	4-19
Chapter 5 – Defining Moral Integrity in the Military		5-1
5.1	Earlier Clarifications on Integrity	5-2
5.2	Advancing an Understanding of Integrity	5-2
5.2.1	Nature of Values and Integrity	5-2
5.2.2	Philosophical Views on Integrity	5-3
5.2.3	Categorising Integrity Definitions	5-4
5.2.3.1	Moral Nature of Standards	5-5
5.2.3.2	External Referent	5-5
5.3	Defining Moral Integrity in the Military	5-6
5.3.1	Methodological Approach	5-6
5.3.2	Findings	5-7
5.3.2.1	Initial Content Analysis	5-7
5.3.2.2	Subsequent Content Analysis	5-7
5.4	Discussion	5-8
5.4.1	Situational Aspects of the Definition	5-8
5.4.2	Implications for Assessing Integrity	5-9
5.4.3	Implications for Research on Integrity	5-10
5.5	Concluding Remarks	5-11
5.6	References	5-11

Chapter 6 – Military Personnel Selection and Technology: Overview	6-1
6.1 Background	6-1
6.1.1 Unique Challenges of Military Personnel Selection	6-1
6.2 Online Personnel Selection Testing	6-2
6.3 Assessment in Military Personnel Selection	6-3
6.3.1 Common Standardised Personnel Assessments	6-3
6.3.1.1 Cognitive Ability Assessments	6-3
6.3.1.2 Personality Assessments	6-4
6.3.1.3 Situational Judgement Tests	6-4
6.3.1.4 Biographical Data	6-4
6.3.1.5 Diversity in Assessments	6-5
6.4 Computer-Based Testing	6-5
6.4.1 Linear on the Fly Test Development	6-6
6.4.2 Computer-Adaptive Tests	6-6
6.4.3 Multi-Stage Adaptive Testing	6-6
6.4.4 Test Item Development	6-6
6.5 Considerations for the Use of Online Personnel Selection Tools	6-7
6.5.1 Integration of Online Selection Tests in a Military Recruiting and Personnel Selection System – The Importance of the Recruitment Website	6-7
6.6 Summary	6-7
6.7 References	6-8
Chapter 7 – Military Personnel Selection and Technology: Evolving Nature of Military Selection	7-1
7.1 Introduction	7-1
7.2 Influence from Inside the Armed Forces	7-2
7.3 Societal Influences	7-4
7.3.1 The ‘War for Talent’	7-4
7.3.2 Increasing Diversity	7-5
7.3.4 The Impact of COVID-19	7-6
7.4 Changes in the Target Populations	7-6
7.5 Resulting Opportunities for Development	7-7
7.5.1 Other Benefits and Opportunities for the Employer	7-8
7.6 Conclusion	7-8
7.7 References	7-9
Chapter 8 – Online Testing Experiences and Benchmarking Survey	8-1
8.1 Introduction	8-1
8.2 Cognitive Tests	8-2
8.2.1 Findings	8-2
8.2.1.1 Test Development	8-3
8.2.1.2 Test Platform	8-3
8.2.1.3 Test Administration	8-4

8.2.1.4	Test Security and Cheating	8-4
8.2.1.5	Planned Improvements in Test Administration	8-6
8.2.2	Summary: Cognitive Tests	8-6
8.3	Personality Tests/Inventories	8-7
8.3.1	Findings	8-7
8.3.1.1	Five-Factor Model	8-7
8.3.1.2	Other Personality Tests	8-8
8.3.2	Summary: Personality Tests	8-8
8.4	Other Tests	8-8
8.4.1	Findings	8-8
8.5	Unproctored Internet Testing	8-9
8.5.1	Concerns About Unproctored Internet Testing	8-9
8.5.2	Benefits of Unproctored Internet Testing	8-9
8.5.3	Experiences with Unproctored Internet Testing	8-10
8.6	Discussion	8-11
8.7	NATO RTG-290 2021 Survey Update	8-12
8.7.1	Findings	8-12
8.7.1.1	Challenges	8-12
8.7.1.2	Solutions	8-12
8.7.1.3	Experiences with Solutions	8-13
8.8	Chapter Summary	8-13
8.9	References	8-14
Chapter 9 – Risks and Risk Mitigation Strategies in Unproctored Internet Testing		9-1
9.1	Introduction	9-1
9.2	Types of Risks	9-1
9.2.1	Psychometric Risk	9-1
9.2.2	Threats to Reliability	9-1
9.2.3	Threats to Validity	9-2
9.2.3.1	Test Security	9-2
9.2.3.2	Authentication	9-3
9.2.3.3	Cheating	9-3
9.2.4	Reputational Risks	9-4
9.3	Risk Mitigation Strategies	9-4
9.3.1	Non-Technical Risk Prevention Strategies	9-4
9.3.2	Technical Risk Prevention Strategies	9-5
9.3.3	Detection Strategies	9-7
9.4	Summary and Conclusion	9-8
9.5	References	9-8
Chapter 10 – Benefits and Costs of Internet-Based Testing		10-1
10.1	Introduction	10-1
10.2	Potential Benefits of Psychological Testing	10-1
10.2.1	Benefits of Computer-Based Testing	10-2

10.2.2	Benefits of Internet-Based Testing	10-2
10.3	Potential Costs of Psychological Testing	10-2
10.3.1	Common Cost-Related Factors	10-3
10.3.1.1	Test Development Costs	10-3
10.3.1.2	Employer and Applicant/Test Taker Costs	10-4
10.3.2	Discretionary Cost-Related Factors	10-4
10.3.2.1	Traditional Proctored Pencil-and-Paper Testing	10-4
10.3.2.2	Computer-Based Testing	10-5
10.3.2.3	Unproctored Testing	10-6
10.3.2.4	Applicant Costs	10-6
10.4	Summary and Conclusion	10-7
10.5	References	10-7

Chapter 11 – Internet-Based Testing on Mobile Devices 11-1

11.1	Introduction	11-1
11.2	Potential Benefits of Testing on Mobile Devices	11-1
11.3	Potential Costs of Testing on Mobile Devices	11-2
11.3.1	Costs for the Organisation	11-2
11.3.1.1	Test Development Costs	11-2
11.3.1.2	Test Maintenance and Version Control Costs	11-3
11.3.1.3	Other Indirect Risks and Associated Costs	11-3
11.3.2	Costs for the Individual	11-3
11.4	Risk and Risk Mitigation	11-4
11.4.1	Potential Disadvantages and Issues Associated with Testing on Mobile Devices	11-4
11.4.1.1	The Very Nature of Being ‘Mobile’	11-4
11.4.1.2	Fairness and Test Equivalence	11-4
11.4.1.3	Equity of Access – People with a Disability	11-5
11.4.1.4	Device Type	11-6
11.4.1.5	Multi-Stage Assessment – Different Tools for Different Jobs?	11-6
11.4.1.6	Applicant Reactions	11-7
11.4.2	Managing Risk in a Mobile Device-Based Selection Context	11-8
11.5	Summary and Conclusion	11-8
11.6	References	11-9

Chapter 12 – Online Testing: Summary and Considerations 12-1

12.1	Summary	12-1
12.2	Considerations	12-1
12.2.1	Online Software Implementation	12-1
12.2.2	Security	12-3
12.2.3	Assessment Management Policy	12-3
12.2.4	Assessment Implementation	12-3
12.2.4.1	The Special Case of Cognitive Ability Assessments	12-4
12.3	Concluding Remarks	12-4
12.4	References	12-4

Chapter 13 – Diversity and Personnel Selection	13-1
13.1 Overview	13-1
13.2 The Concept of Diversity	13-1
13.3 Current State of Diversity in Military Personnel Selection	13-2
13.4 Implications of Diversity for Personnel Selection	13-2
13.5 Fairness Considerations in Personnel Selection	13-2
13.5.1 Test Development	13-3
13.5.2 Test Administration	13-3
13.5.3 Test Score Interpretation	13-5
13.6 Neurodiversity: An Emerging Consideration	13-6
13.6.1 Neurodiversity in Organisations	13-6
13.6.2 Neurodiversity in Defence and Related Sectors	13-6
13.6.2.1 Israeli Defence Forces (IDF)	13-6
13.6.2.2 Australian Defence Organisation (ADO) and DXC Technology	13-7
13.6.2.3 The United Kingdom	13-7
13.6.3 Implications of Neurodiversity in Personnel Selection	13-7
13.7 Summary	13-8
13.8 References	13-8
Chapter 14 – Advances in Military Personnel Selection: Next Steps	14-1
Annex A – Online Testing: Standardised Terminology	A-1
Annex B – Online Testing Benchmarking Survey (2018)	B-1
Annex C – Online Testing Benchmarking Survey (2021 Update)	C-1

List of Figures

Figure		Page
Figure 8-1	Overview of Topics for the Survey	8-1
Figure 12-1	Considerations When Implementing Online Assessments	12-2

List of Tables

Table		Page
Table 3-1	RFI Questions on Integrity Testing	3-1
Table 3-2	Overview Measuring Integrity During Selection	3-2
Table 3-3	Overview of Methods Used to Assess Integrity Information in the Selection Decision	3-3
Table 4-1	Commercial Integrity Tests Currently in Print	4-3
Table 4-2	Integrity Tests Used in Scandinavian Organizations	4-8
Table 5-1	A Categorisation of Existing Integrity Definitions/ Perspectives by Referent and Nature of Values	5-4
Table 5-2	Results of Content Analysis of High-Level Codes from NATO Military Institutions	5-8
Table 9-1	Summary of Risks, Risk Prevention and Detection Strategies	9-8
Table 10-1	Cost-Related Factors for Different Forms of Tests	10-3
Table 13-1	R&D Considerations Based on the Elements of Universal Design	13-4

Acknowledgements

We acknowledge Dr. Meghan Fitzpatrick, Mr. Geoff Gallas, Mr. Maxime Grenier, and Dr. Catherine Smith for their editorial contributions in reviewing, revising, and/or proof-reading draft versions of various chapters in this report. Special thanks to Mr. Geoff Gallas for taking on a significantly higher load of this arduous task.

We also acknowledge Capt Emmanuel Frere (FRA), Dr. Adele Luta (USA), Dr. Rabiah Muhamad (USA), Ms. Elizabeth Salmon (USA), and Dr. Heather Wolters (USA) for their participation in some activities of this RTG.

HFM-290 Membership List

CO-CHAIRS

Dr. Wendy DARR*
National Defence
CANADA
Email: wendy.darr@gmail.com

Dr. Elliot LOH*
National Defence
CANADA
Email: elliott.loh@forces.gc.ca

MEMBERS

Mr. Jimmi ANDREASEN*
Danish Defence Personnel Organisation
DENMARK
Email: jimmi.andreasen@mil.dk

Mr. Jan Oddvar HEIMDAL*
Norwegian Defence University College
NORWAY
Email: janoheimdal@gmail.com

Dr. Françoise BERTRAND
Belgium Defence
BELGIUM
Email: francoise.bertrand@mil.be

Mr. Stefan HOFER*
Federal Office of Bundeswehr Personnel
Management
GERMANY
Email: stefan1hofer@bundeswehr.org

Dr. Sybille DUNKER*
Federal Office of Bundeswehr Personnel
Management
GERMANY
Email: sibylledunker@bundeswehr.org

LCol Soo MYUNG OWENS*
Canadian Forces Recruiting
CANADA
Email: soo.owens@forces.gc.ca

Dr. Meghan FITZPATRICK*
Centre for Operational Research and Analysis
Defence Research and Development Canada
CANADA
Email: meghan.fitzpatrick@forces.gc.ca

Col Marek NOVY (ret)
Czech National Police
CZECH REPUBLIC
Email: marek.novy@pcr.cz

Dr. Maria FORS BRANDEBO*
Swedish Defense University
SWEDEN
Email: maria.forsbrandebo@fhs.se

Mr. Kai NYMAN*
Finnish Defence Research Agency
FINLAND
Email: kai.k.nyman@mil.fi

Mr. Geoff GALLAS*
Joint Health Command
AUSTRALIA
Email: geoff.gallas1@defence.gov.au

Dr. Tom H. SKOGLUND*
Norwegian Armed Forces
NORWAY
Email: tskoglund@mil.no

Mr. André HAZEL
Army Recruiting and Training Division
UNITED KINGDOM
Email: andre.hazel100@mod.uk

Dr. Charlene STOKES-SCHWARTZ
The MITRE Corporation
UNITED STATES
Email: cstokes@mitre.org

* Contributing Author

Dr. Marjolein VERBOOM*
Department of Defence
NETHERLANDS
Email: mj.verboom@mindef.nl

Dr. Markéta ZAJÍCOVÁ*
Military University Hospital Prague
CZECH REPUBLIC
Email: marketa.zajicova@uvm.cz

Mr. Gerhard WOLGERS*
Swedish Armed Forces
SWEDEN
Email: gerhard.wolgers@mil.se

ADDITIONAL CONTRIBUTORS

Ms. Kelly BATTOCCHIO*
Department of Defence
AUSTRALIA

Ms. Janne Hellerup NIELSEN*
Danish Armed Forces
DENMARK

Ms. Annemie DEFRANC*
Belgian Defence
BELGIUM

Dr. Catherine SMITH*
British Army Recruiting and Initial Training
Command (ARITC)
UNITED KINGDOM

Mr. Maxime GRENIER*
Department of National Defence
CANADA

Maj Samuel WILLIAMS
New Zealand Defence Force
NEW ZEALAND

Maj Joel MAJER
New Zealand Defence Force
NEW ZEALAND

PANEL/GROUP MENTOR

Ms. Eugenia KALANTZIS
Department of National Defence
Canada
Email: Eugenia.kalantzis@forces.gc.ca



Advances in Military Personnel Selection

(STO-TR-HFM-290)

Executive Summary

Human capital and human resource systems that produce human capital (e.g., personnel selection, training) are important elements of an organization's ability to adapt to change. Yet, such processes must also remain responsive to change in order to remain efficient and effective. In this report, we explore topics within the umbrella of personnel selection, to identify current trends, challenges, and developments in the field. We focus on three specific areas that have significance in today's defence context: integrity, online/computerised testing, and diversity.

Given the breadth of each topic area, three sub-groups of participating nations with an interest in a particular topic were formed within this Research Task Group. Each sub-group identified and outlined specific areas that were explored by the group, based on an understanding of current needs, challenges, or requirements of their respective organizations. The resulting output for each topic is documented in separate sections of this report.

With respect to integrity, this report provides insight into the history of integrity testing in the military; examines the current state of integrity testing practice and research across participating nations; summarizes commercial integrity tests and emerging approaches to assessing integrity; and provides a framework for defining integrity in the military for the purposes of its assessment in military personnel selection. On the topic of online/computerised testing, this report examines the selection testing practices of participating nations, in addition to exploring psychosocial factors, testing security challenges, options for online test administration, and the benefits to personnel selection that can result from technology implementation. In the area of diversity, this report explores the concept of diversity; outlines fairness and inclusivity considerations in personnel selection practices; and explores neurodiversity initiatives within defence.

All three topics (i.e., integrity, online/computerised testing, and diversity) are especially relevant to the defence context. The selection of individuals with the propensity to act with integrity is important to military personnel selection practices because public confidence and trust are paramount to the effectiveness of the Armed Forces. To ensure competitive advantage in the recruitment of top talent, military personnel selection testing practices must also appeal to technologically savvy labour markets, while embodying the principles and fairness and inclusivity as these markets become increasingly heterogeneous in demographic (e.g., ethnicity) and non-demographic (e.g., language) factors. In summary, this report draws upon the current state of practice and research to raise awareness of and provide guidance on topics that deserve consideration in military personnel selection.

Progrès dans la sélection du personnel militaire

(STO-TR-HFM-290)

Synthèse

Le capital humain et les systèmes de ressources humaines qui produisent le capital humain (par exemple, la sélection du personnel et la formation) sont des éléments importants de la capacité d'une organisation à s'adapter au changement. Cependant, ces processus doivent également réagir au changement afin de rester efficaces et rentables. Dans le présent rapport, nous étudions des sujets se rattachant à la sélection du personnel, pour identifier les tendances actuelles, les défis et les évolutions dans ce domaine. Nous nous concentrons sur trois domaines spécifiques qui ont de l'importance dans le contexte actuel de la défense : l'intégrité, les tests en ligne/informatisés et la diversité.

Étant donné l'étendue de chaque domaine, trois sous-groupes de pays participants éprouvant de l'intérêt pour un sujet particulier ont été constitués au sein du présent groupe de recherche. Chaque sous-groupe a identifié et décrit les domaines spécifiques qui l'intéressaient, en fonction de la compréhension des besoins, défis ou exigences actuels des organisations concernées. Le résultat est documenté pour chaque sujet dans une partie distincte du présent rapport.

Au sujet de l'intégrité, le présent rapport expose l'histoire des tests d'intégrité dans le domaine militaire, examine l'état actuel de la pratique et des recherches sur les tests d'intégrité dans les pays participants, résume les tests commerciaux d'intégrité et les approches émergentes d'évaluation de cette dernière et fournit un cadre de définition de l'intégrité dans le domaine militaire afin de permettre son évaluation pendant la sélection du personnel militaire. À propos des tests en ligne/informatisés, le présent rapport examine les tests de sélection pratiqués par les pays participants, en plus d'explorer les facteurs psychosociologiques et de tester les défis de sûreté, les options d'administration des tests en ligne et les avantages pour la sélection du personnel qui peuvent découler de l'application de la technologie. Dans le domaine de la diversité, le présent rapport étudie les concepts de diversité, décrit les considérations d'équité et d'inclusivité dans les pratiques de sélection du personnel et explore les initiatives de neurodiversité au sein de la défense.

Les trois sujets (intégrité, tests en ligne/informatisés et diversité) sont particulièrement pertinents dans le contexte de la défense. La sélection d'individus ayant la propension à agir avec intégrité est importante dans les pratiques de sélection du personnel militaire, parce que la confiance du public est primordiale pour l'efficacité des forces armées. Afin de garantir l'avantage concurrentiel dans le recrutement d'éléments de valeur, les pratiques de test de sélection du personnel militaire doivent également faire appel aux marchés du travail exploitant les nouvelles technologies et incarner les principes, l'équité et l'inclusivité, car ces marchés sont de plus en plus hétérogènes sur le plan démographique (par exemple, l'origine ethnique) et non démographique (par exemple, la langue). En résumé, le présent rapport s'appuie sur l'état actuel de la pratique et des recherches pour sensibiliser le lecteur et lui prodiguer des conseils sur des sujets qui méritent d'être pris en considération pendant la sélection du personnel militaire.

Chapter 1 – AN INTRODUCTION TO ADVANCES IN MILITARY PERSONNEL SELECTION

Wendy Darr and Elliot Loh
Department of National Defence
CANADA

1.1 BACKGROUND

It is well-known among management and organisational scholars that organisations must respond and adapt to an ever-changing environment in order to remain effective or to maintain competitive advantage. Military organisations are no exception; they are complex adaptive systems that evolve in response to their environment to survive [22]. Changing national and/or international social, economic, and/or geo-political landscapes are environmental factors to which military organisations must respond. These responses are often reflected in a nation’s defence priorities and strategies. For example, the Australian Defence Force’s (ADF) 2016 Strategic Outlook highlights its proposed response over the next few decades to environmental uncertainty, which includes security challenges and economic and political shifts in power in the Indo-Pacific region [1]. In doing so, it ensures that the future ADF remains “potent, agile and ready to respond” (p. 2). Similarly, Canada’s defence policy reflects its long-term vision and response to changes in the world [12].

Organisational adaptability, defined as the extent to which an organisation responds to and adapts to changes in its environment [20], is dependent on many organisational elements, one of which is its people. People are the basic element of all organisations, and the extent to which they are positioned to interpret and respond to change is an important determinant of organisational adaptability [21]. In Ployhart and Turner’s [20] model of organisational adaptability, human capital resources is one of four components at the microfoundation of organisational adaptability; human capital resources, in turn, emerge from individual knowledge, skills, abilities, and other attributes (KSAOs). Another component is the organisation’s human resources systems (e.g., selection, training, compensation) that comprise the policies and processes that facilitate the emergence of human capital resources from individuals’ KSAOs. Of these various systems, personnel selection processes are vital to the emergence of human capital resources [19]. Not surprisingly, in support of the United States Army’s transformation to its Future Force, significant efforts were dedicated to initiatives pertaining to the selection of future entry-level soldiers to ensure that the Army selects and classifies those with the KSAOs required to successfully perform in a transformed Army [10].

In this NATO Research Task Group (RTG-290), efforts were focussed on topics pertinent to military personnel selection. Although prior NATO STO activities have examined personnel selection, those efforts were directed at the selection practices of specific sub-groups such as Special Operations Forces (HFM-17 [RTG]; 2008 –2012) and Officers (HFM-032[RWS]; 1999 – 2001). Their resulting reports do not capture emerging challenges, trends, or developments in the field of personnel selection as they relate to current defence priorities. In this RTG, we examined three topics of relevance to military personnel selection in today’s context: integrity-related assessment online/computerised testing, and diversity. At the outset, the objectives of this RTG were to:

- Identify military personnel selection and assessment issues/challenges experienced by nations, especially in areas pertinent to integrity-related assessments, online/computerised testing, and diversity;
- Identify current approaches/strategies used in nations to the deal with the identified issues/challenges; and

- Identify best practices and emerging trends in selection and/or assessment tools, processes, or strategies as they pertain to integrity-related assessments, online/computerised testing, and diversity in military selection.

Given the breadth of each topic area, three sub-groups were formed within this RTG, based on participating member interests. The resulting work of each sub-group is documented in separate sections of this report. Below, we provide a brief overview of the relevance of each topic area to the field of personnel selection, as well as their significance to the defence priorities and strategies of many nations.

1.2 RELEVANCE OF TOPICS

A scientometric study that used text mining software to comb patent and publication repositories for subject areas published in the personnel selection domain between 2007 and 2017 revealed the three topics of interest to this RTG (i.e., integrity-related assessment, online/computerized testing, and diversity) to have research and development (R&D) momentum worthy of interest [3]. Using an R&D momentum indicator (based on publication count and publication growth), Culhane and Soles [3] grouped subject areas into four quadrants (established, hot, emerging, and brand new or declining). Subject areas pertaining to online testing (e.g., online tests, unproctored testing, computer adaptive testing, gamification) were grouped in the *emerging topics* quadrant. In other words, the number of publications in this area increased over the years, with a larger concentration in the area of online screening as well as faking or cheating on tests. Also included in the *emerging topics* quadrant were areas pertinent to diversity (e.g., discrimination, gender differences in selection or testing contexts). Publications in this area pertained to differential treatment of, or the influence of, applicant attributes such as age, ethnicity, gender, or appearance on selection outcomes. Finally, the subject of integrity was grouped in the *hot topics* quadrant, indicating a rapid number and growth of publications over the years. Publications on this subject pertained to validity examinations of mostly personality-based integrity tests in predicting counterproductive work outcomes. Together, these findings highlight a growing research interest in the three aforementioned areas. Below, we further illustrate the relevance of each area to the priorities of defence organisations, and briefly describe the outputs produced by this RTG as reflected in various chapters of this report.

1.2.1 Integrity-Related Assessments

Military organisations are often described as having a dual responsibility, internally (to their personnel) and externally (to the public; [23]). As such, public and personnel confidence and trust are paramount to the effectiveness of the Armed Forces. However, this trust can be easily eroded with reports of corruption and misconduct in military organisations. A survey of NATO member and partner defence establishments undertaken by Transparency International, a United Kingdom Security and Defence program, found vulnerability to corruption to be quite high across 32 NATO operations [11]. The report also indicated that only four out of 22 nations addressed corruption in their respective doctrines. Sexual misconduct and discrimination are also issues that plague the militaries of several nations (e.g., Refs. [9], [4], [14]). Consequently, a NATO initiative called *Building Integrity* is presently directed at identifying good governance policies and structures to improve transparency and accountability [15]. A related output called the *Allied Command Operations (ACO) Handbook* about building integrity in operations discusses integrity at various levels, such as political, organisational, and personal [15]. For instance, at the personal level, the Handbook outlines the role of military leadership in establishing accountability by leading by example; however, being macro in focus, the book does not discuss personnel selection requirements (i.e., the need to select leaders most likely to lead by example). To yield a critical mass of personnel who, for example, demonstrate integrity in the workplace, requires the selection of individuals with the necessary KSAOs that are linked to integrity (cf., Ref. [19]).

Aligned with efforts to build integrity in defence, the output yielded by this RTG provides a brief look into the history of integrity testing in the military and examines the current state of integrity testing practice and research across participating nations. It also provides insight into various commercially used and emerging approaches to assessing integrity. Finally, the report sheds light on varied conceptions of integrity, and provides a framework for defining integrity to facilitate consistency in how it is conceptualised and should be assessed in personnel selection. Part I of this report contains Chapters 2 through 5 that focus on the topic of integrity-related assessments.

1.2.2 Online/Computerised Testing

The relevance of online testing to defence is reflected in an increasing focus on technological advances in the capability development plans of the European Defence Agency [5]. The modernisation of its business processes is also a defence priority for Canada [12]. Relatedly, with respect to recruitment and selection, the Canadian Armed Forces has focussed on e-recruiting techniques and strategies that pivot on advanced technologies [2]. This RTG focussed on online testing and the application of technological advances in personnel selection testing and processes. The selection practices of organisations can influence organisational brand image, which in turn, influences competitive advantage [19]. For example, high-tech companies will often leverage innovative technologies in their recruitment and selection processes because it portrays an image that is consistent with their brand, and consequently attracts top talent [19]. Given that future labour markets will comprise an increasing proportion of Generation Z individuals who have been exposed to the internet since birth, and are extremely familiar and dependent on technology (e.g., Refs. [6]; [16]), technology will play an important role in the personnel selection practices of organisations in attracting and selecting this generation.

The incorporation of advanced technologies for test administration also affords far greater flexibility in how assessments are administered, thereby opening opportunities for both developing innovative personnel selection approaches and revamping traditional ones. The adoption and integration of technology to advance personnel selection processes, however, requires the careful consideration of a range of factors that could impact the psychometric efficacy of the assessments. Therefore, in addition to exploring the online selection testing practices of military nations participating in this RTG, a range of pervasive issues including test security, test-taker authentication, cheating in unproctored testing conditions, as well as psychosocial issues and technology options for online test administration are also reviewed. Potential resource requirements for the implementation of advanced online assessment technologies are also detailed in relation to possible benefits afforded by the technology implementation. Altogether, these chapters provide a comprehensive overview of advanced online testing approaches available for implementation in military personnel selection. Part II (Chapters 6 through 12) is dedicated to topics pertaining to online/computerized testing.

1.2.3 Diversity

Diversity, which pertains to real or perceived heterogeneity resulting from demographic (e.g., race, age) and non-demographic (e.g., religion) factors, remains a long-standing and complex human resources management issue for military leaders [18]. Although some nations approach diversity as a response to legislation in meeting certain requirements [7], there is an increasing recognition of the value of diversity to military effectiveness. In highlighting the imbalance of women in the military, for example, Hörst, Mattox, and Groenendaal [8] discussed the value of having more women in military missions to bolster public trust and local community engagement. The impetus for the United Kingdom's diversity and inclusion strategy is also articulated in terms of operational effectiveness [13]. In a policy brief of the German Marshall Fund of the United States, which aims to strengthen trans-Atlantic cooperation on global challenges, all members and partners of the European Union and NATO are called to increase attention, pressure, and accountability to encourage diversity and inclusion at all military levels [8].

With increased global migration, diversity within countries has increased over the decades, as suggested by a survey of 27 nations, including North American, European, and Asia-Pacific countries [17]. The labour markets of North American, European, and Asia-Pacific countries will continue to become more heterogeneous in terms of their demographic (e.g., ethnicity, sex, age) and non-demographic (e.g., language, religion, sexual orientation) composition (cf., Ref. [17]). Given this reality, entry-level military selection processes must be positioned to facilitate, rather than hinder, a military's diversity efforts. This RTG explored the concept of diversity, examining specific considerations that must be given to personnel selection practices to ensure fairness and inclusivity to all applicant groups. It also explored the concept of neurodiversity, an emerging area within the field with further implications for personnel selection practice within defence. Chapter 13 in Part III of this report discusses issues that pertain to diversity and personnel selection.

1.3 CLOSING REMARKS

In its coverage of three areas (integrity, online testing, and diversity), this report represents a comprehensive resource on topics that are highly relevant to personnel selection, particularly in the defence context. Drawing upon the current state of practice and research, the report is intended to raise awareness or provide guidance on issues that must be given consideration when designing selection systems to incorporate integrity-related assessments, online/computerised testing, or to ensure fairness for diverse groups.

1.4 REFERENCES

- [1] Australian Defence Force. (2016). 2016 Defence White Paper: Strategic outlook. Australian Defence Force. <https://www.defence.gov.au/whitepaper/>
- [2] Canadian Armed Forces (2014). Strategic level guidance: An evolving landscape.
- [3] Culhane, M., and Soles, J. (2018). Emerging trends in personnel selection: A scientometric study (NRC-NSL Project Number: MC18-01). National Research Council Canada.
- [4] Deschamps, M. (2015). External review into sexual misconduct and sexual harassment in the Canadian Armed Forces. <https://www.canada.ca/en/department-national-defence/corporate/reports-publications/sexual-misbehaviour/external-review-2015.html>
- [5] European Defence Agency (2019). The main architect of EU defence capability priorities. Retrieved from <https://www.eda.europa.eu/webzine/issue17/key-achievements/the-main-architect-of-eu-defence-capability-priorities/>
- [6] Goh, E., and Lee, C. (2018). A workforce to be reckoned with: The emerging pivotal Generation Z hospitality workforce. *International Journal of Hospitality Management*, 73, 20-28.
- [7] Holden, N.J. and Davis, K. (2006). Harassment in the military: Cross-national comparisons. In F.C. Pinch, A.T. MacIntyre, P. Browne, and A.C. Okros (Eds.). *Challenge and Change in the Military: Gender and Diversity issues*. Canadian Defence Academy Press. Retrieved 1 Feb 2021 from <https://wiisglobal.org/wp-content/uploads/2013/05/Challenge-and-Change-in-the-Military-Gender-and-Diversity-Issues2.pdf>
- [8] Hörst, C., Mattox, G.A., and Groenendaal, L. (2018). Raising EU and NATO Effectiveness: The impact of diverse boots on the ground. Policy Brief. The German Marshall Fund of the United States. <https://www.gmfus.org/publications/raising-eu-and-nato-effectiveness-impact-diverse-boots-ground>

- [9] Judicial Proceedings Panel Subcommittee (2017). Report on sexual assault investigations in the military. Retrieved 1 Feb 2021 from https://media.dothanfirst.com/nxsglobal/dothanfirst/document_dev/2019/01/08/JPP_Investigations_Final_Feb%202017_1546991720274_66889432_ver1.0.pdf
- [10] Knapp, D.J., Sager, C.E., and Tremble, T.T. (2005). Development of experimental Army enlisted personnel selection and classification tests and job performance criteria (Technical Report 1168). U.S. Army Research Institute for the Behavioral and Social Sciences. <https://apps.dtic.mil/sti/citations/ASA438314>
- [11] Maclachlan, K., and Hurd, H. (2015). How corruption undermines NATO operations. Retrieved 1 Feb 2021, from <https://www.defenseone.com/ideas/2015/12/how-corruption-undermines-nato-operations/124148/>
- [12] Minister of National Defence. (2017). Strong, secure, engaged: Canada's defence policy. <https://www.canada.ca/en/department-national-defence/corporate/policies-standards/canada-defence-policy.html>
- [13] Ministry of Defence. (2018). A force for inclusion: Defence diversity and inclusion strategy 2018 – 2030. <https://www.gov.uk/government/publications/defence-diversity-and-inclusion-strategy-2018-to-2030-a-force-for-inclusion>
- [14] Moulson, G. (2020). Germany apologizes for past military anti-gay discrimination. Retrieved 1 Feb 2021 from <https://www.thestar.com/news/world/europe/2020/09/17/germany-apologizes-for-past-military-anti-gay-discrimination.html>
- [15] NATO (2020). Building integrity in operations: ACO handbook. https://www.nato.int/nato_static_fl2014/assets/pdf/2020/6/pdf/200608-bi-handbook.PDF
- [16] Ozkan, M., and Solmaz, B. (2015). The changing face of the employees-generation Z and their perceptions of work. *Procedia Economics and Finance*, 26, 476-482. doi: 10.1016/S2212-5671(15)00876-X
- [17] Pew Research Center (2019). A changing world: Global views on diversity, gender equality, and family life, and the importance of religion. <https://www.pewresearch.org/global/2019/04/22/a-changing-world-global-views-on-diversity-gender-equality-family-life-and-the-importance-of-religion/>
- [18] Pinch, F. (2006). An introduction to challenge and change in the military: Gender and diversity issues. In F.C. Pinch, A.T. MacIntyre, P. Browne, and A.C. Okros (Eds.). *Challenge and Change in the Military: Gender and Diversity Issues*. Canadian Defence Academy Press. Retrieved 1 Feb 2021 from <https://wiisglobal.org/wp-content/uploads/2013/05/Challenge-and-Change-in-the-Military-Gender-and-Diversity-Issues2.pdf>
- [19] Ployhart, R.E. (2012). Personnel selection and the competitive advantage of firms. In G.P. Hodgkinson and J.K. Ford (Eds.). *International Review of Industrial and Organizational Psychology*, 153-195. Wiley Blackwell.
- [20] Ployhart, R.E., and Turner, S.F. (2014). Organizational adaptability. In D. Chan (Ed.), *Individual adaptability to changes at Work: New directions in research*, 73-91. Routledge/Taylor & Francis Group.

- [21] Rybakov, L. (2001). Environment and complexity of organizations. *Emergence: Complexity and Organization*, 3(4). doi:10.emerg/10.17357.afead99adb24518903ba336f63aa25de
- [22] Surace, A. (2019). Complexity and leadership: The case of a military organization. *International Journal of Organizational Analysis*, 27(5), 1522-1541. doi.org/10.1108/IJOA-12-2018-1614
- [23] Walzer, M. (1981). Two kinds of military responsibility. *Parameters: Journal of the US Army War College*, 11(1), 42-26.

Chapter 2 – HISTORY OF INTEGRITY TESTING

Meghan Fitzpatrick

Defence Research and Development Canada
CANADA

2.1 INTRODUCTION

“Man is the first weapon of battle. Let us study the soldier for it is he who brings reality to it.”

Ardant Du Picq ([23], p. 44)

Designed to predict “counterproductive workplace behaviour” ([72], p. 57) integrity testing is the latest stage in the long and rich history of personality testing in the military.¹ A controversial field, it has nonetheless become an essential part of life in the armed forces over the past hundred years. This chapter will provide a historical overview of developments in the field that have contributed to the current integrity-based orientation of testing. For the purposes of this piece, personality testing will be defined as psychological testing methods that seek to measure specific traits or characteristics as a means to predict future behaviour and/or performance ([68], pp. 71-83).

Personality testing has had a far-reaching impact on militaries around the world. In light of this consideration, this chapter will examine three countries, namely: Canada, the United Kingdom, and the United States. The choice of countries reflects both the accessibility of source materials in English and the significant influence that each nation has had internationally, particularly the United States. Reflecting on the past century, it seeks to answer three fundamental questions. First, what factors motivated and contributed to military investment in personality testing? Second, what results were achieved and how did testing practices evolve over time? Finally, and most importantly, what historical lessons can the defence scientific community derive that can be applied to future development?

To best answer these questions, the chapter will explore the origins of the military’s interest in psychometric testing during the First World War (1914 – 1918) and what factors contributed to the dramatic expansion of psychological screening during the Second World War (1939 – 1945). This will be followed with a look at the reasons why investment in personality testing continued to increase during the Cold War, despite the questionable results achieved over the preceding decades. In addition, it will examine the research conducted at the height of the Cold War, when personality testing became a better integrated element of the military personnel system. This includes dissecting the rationale behind the negative orientation of screening methods and the socio-cultural considerations that informed this theoretical focus. Finally, I will reflect on how the failure to design accurate and consistent methods to screen out maladaptive traits and the development of new theories of personality ultimately drove a reorientation of testing methods.

2.2 PERSONALITY TESTING (1914 – 1945)

Modern personality testing has its origins in the late nineteenth century, when scholars first began to devise “explicit experimental procedures for measuring psychological attributes” ([13], p. 4). Much of this early work

¹ Please note that integrity testing can be defined in a variety of ways and the conception used here reflects that used by the remaining authors in this report.

centred on assessing school children in growing national education systems and workers in increasingly complex industrial enterprises [52]. Military forces only began to adopt similar methods decades later, as mobilisation for the First World War began [13]. The grinding reality of attritional warfare necessitated the rapid recruitment of an unprecedented number of untrained civilians. Moreover, these same enlistees had to be allocated to appropriate roles across the armed forces as quickly as possible. The scale of mobilisation efforts was vast. The United States alone deployed close to three million service personnel overseas [41]. Meanwhile, Britain and its Dominions raised nearly four million soldiers and labourers from 1914 to 1918 [53]. Inefficient use of valuable manpower was both financially wasteful and could result in costly defeat on the battlefield. As University of Virginia historian Erik Lunstrum has so cogently argued, psychological testing “helped to meet a daunting logistical challenge” [45].

The United States military developed some of the earliest screening and selection instruments [36]. Total war stimulated the rapid expansion of the existing recruitment apparatus and the development of innovative psychometric tools, including the first self-report personality inventory. Commissioned by the United States Army, the Woodsworth Personal Data Sheet included 116 yes or no questions intended to measure emotional stability [32]. Questions ranged from the mundane, such as “Do you usually feel in good health?” to the pointedly intrusive like, “Do you drink a fifth of whiskey a day?” and “Do you wet the bed at night?” [77]. Academic psychologist and inventor, Professor Robert S. Woodsworth developed the test based on interviews with clinicians and through examining case studies of psychologically unstable individuals. While the United States Army never had the chance to deploy the test operationally, the data sheet soon proved popular. Easy to reproduce and administer, it is still widely considered to be the predecessor of today’s personality questionnaires and inventories [13]. Its introduction also coincided with pioneering developments in projective testing, where subjects are asked to respond to ambiguous stimuli to detect personality characteristics and to identify abnormal psychopathology. The most famous of these tests includes both the Rorschach Ink Blot and the Thematic Apperception Test (TAT) [13].

In spite of the demand for efficient testing methods, the introduction of screening instruments like the data sheet was not without dispute. Throughout the first half of the twentieth century, military circles remained highly sceptical of the new *human sciences* [30], [39]. On both sides of the Atlantic, the lay public also continued to associate psychologists and psychiatrists with the indignity and stigma of asylum medicine [66]. It is important to note that incorporating scientific theory into the recruitment process also directly challenged traditional martial values and entrenched ideas about the characteristics that defined the warrior class at the turn of the twentieth century [11], [31]. The introduction of screening and selection represented a logical response to manage the incoming flow of manpower. Over time, it also appeared to offer a preventative measure to quell the swelling number of shell-shock patients filling regimental aid posts and frontline field dressing stations [67]. As early as December 1914, the British Expeditionary Force reported that 7% to 10% of its officers and 3% to 4% of other ranks were suffering from nervous illnesses [67]. Having said that, this decision should not be seen as signalling a fundamental shift in military attitudes towards the mental health professions nor the methods they employed. When the armistice was signed in 1918, Canadian, British, and American military officials discarded newly adopted screening measures as an anomaly of total war and largely returned to pre-1914 methods of assessment [23], [43].

In direct contrast, industry and the private sector recognised the transformative potential of these methods and helped to spur on the advancement of industrial and organisational psychology during the interwar period (1919 – 1939). Like the armed forces, businesses required efficient methods with which to process hundreds, if not thousands, of new employees. The organisational complexity and size of businesses grew exponentially in the first half of the twentieth century [64]. Personality testing and psychometric methods fit neatly into the zeitgeist of scientific management and efficiency that rose to prominence during this period with organisational theorists like Australian psychologist Elton Mayo (1880 – 1949) [32].

The business community also presented a rich marketplace for personality inventories and any initiative that claimed to offer executives a new way to mobilise and manage the workforce. The Humm-Wadsworth Temperament Scale (HWTS) provides a notable example. Developed in the mid-1930s, the test's creators advertised the HWTS as an accurate predictor of counterproductive workplace behaviour and quickly found a receptive audience. A little less than a decade later, a *Time* magazine article entitled "Pegs that Fit" claimed that over 2,000,000 American workers had taken the HWTS ([32], p. 176). Similar to other inventories produced at the time, the HWTS focused on identifying the "negative and maladaptive" ([32], p. 167) components of personality that hindered effective integration into the workplace.

The advent of the Second World War prompted renewed military interest in psychological screening and selection, which included personality testing. The financial cost of compensating and caring for thousands of traumatised First World War veterans had served as a salutary financial lesson and an incentive to invest in preventative measures. Globally, pensions and medical care from the Great War represented a heavy cost to the former belligerents. Nearly a decade after hostilities had ended, close to 9,000 Canadians continued to claim compensation for psychological problems [22], and on the eve of the Second World War, roughly 120,000 former British service personnel were "still in receipt of pensions or had received final awards for primary psychiatric disability" ([2], p. 108). In the early 1940s, the United States Government estimated that on average, "every case of neuropsychiatric disability among veterans of the last war [would], during the man's life-time, cost...\$30,000, and that about a billion dollars [had] already been spent" ([76], p. 301).

Informed by these lessons, Canada, the United Kingdom, and the United States all poured resources into screening and selection. Armed forces in all three countries saw personality as a vital component of stability, motivation, and capacity to adapt to military life [19]. Consequently, personality testing represented an essential component of the screening process. Early in the war, Personnel Selection Officers (PSOs) primarily used structured interviews to assess new recruits [54]. The interview provided officials with the opportunity to probe candidates about everything from education to occupational history, health, and hobbies. As Canadian military psychologist J. W. Howard [37] explained, "what is wanted...is a useful conception of the subject as a functioning person and not simply as an aggregate of adjectival description, all-or-none summation of factual characteristics or the noting of the presence of neurotic symptoms." While the interview had its merits, the accuracy of its findings relied heavily on "the skill and the acumen" [37] of the individual PSO or clinician involved. Furthermore, detailed examinations like this lost their attraction as the demand for manpower skyrocketed. Necessity dictated the development of more paper-and-pencil questionnaires, personality inventories, and other psychometric methods [20]; [54]. Quantifiable, reproducible, and easily administered, these tools were well suited to the increasingly industrialised nature of conflict [58].

The popularity of personality testing reached unprecedented heights during the Second World War. Psychologists developed tests to identify and measure everything from grit to combat effectiveness and to better predict those who would break down on the battlefield. But subsequent reviews of the screening process suggest that reality could not measure up to these extravagant claims. For instance, the Canadian Army released more than 200,000 men as unfit for service for psychiatric reasons throughout the Second World War [21]. However, a 1943 follow-up study of 22,000 soldiers discharged as psychoneurotic indicated that "almost all had found civilian employment, with 43 percent reporting that they now had a 'better job' than before enlisting [21]." The American Selective Service System produced similarly troubling and inconsistent results. In a "Study of Neuropsychiatric Rejectees," researchers Egan, Jackson and Eanes [29] tracked the careers of over 2,000 men rejected by the "Selective Service System on psychiatric grounds but later inducted into [the armed forces]." Based on their findings, they estimated that the United States Army had unnecessarily dismissed roughly 2,000,000 men from military service between 1941 and 1945.

Studies commissioned in the early Cold War further underline the problems involved in accurately measuring and quantifying human personality. In a report for Canada's Defence Research Board, consultant psychiatrist Franklin Chalke bluntly confirmed what many already knew, namely that "existing methods of screening...could at best detect 50 per cent of those who would fail and would reject an equal number of those who would become satisfactory soldiers." Despite sustained effort, he concluded that the armed forces were "little further ahead than...at the beginning of World War II in knowing what factors are important in assessing a recruit [21]."

Chalke's voice was one of many in a growing chorus. In 1950, the Commonwealth Advisory Committee on Defence Science (CACDS) concluded that despite the extent of wartime investment, personality testing remained "an immense subject which bristles with theoretical and practical problems...many of which lie well beyond the limited field with which Service psychologists can concern themselves [20]." Famous American clinician Karl Menninger entirely dismissed wartime testing in his 1948 book, *Psychiatry in a Troubled World* [48]. There he argued that "there is no doubt that the method by which the psychiatrist made his personality assessment...at induction had many flaws. Men were rejected who could have given good service, while other men were accepted who should have been rejected." He went on to further conclude that [48]:

...our experience with combat taught us that no selection methods could have picked out the men who would fail. In other words, if a man succeeded in maintaining his mental equilibrium through the training period, shipment overseas and training on foreign soil, there was no means by which unsuccessful combat service could have been forecast.

2.3 COLD WAR RESEARCH (1945 – 1970s)

Postwar backlash was strong and vocal. Nonetheless, many nations remained committed to refining existing personality assessment tools and funding new initiatives. There are several underlying reasons why personality testing and psychometric methods remained appealing. In the event of another world war, these techniques still offered the best chance of processing new waves of recruits quickly and efficiently. Both the United Kingdom and the United States also continued to impose forms of national service or conscription throughout the 1950s and 1960s [75]. As the Cold War deepened, the technical nature of modern war dictated the need to recruit increasingly agile and adaptable candidates. Finally, there were numerous financial incentives to pre-emptively identify recruits who might perform poorly or exhibit counterproductive behaviour.

Canada's recruitment woes during the Korean War (1950 – 1953) provide one such example. Throughout this period, the Canadian Army discharged nearly 7,000 soldiers, after an average of only eighteen months, for psychological reasons [16]. A contemporary report on manpower conservation estimated that the "cost of paying, keeping, feeding, training, transporting, disciplining and treating [these men]" neared \$35,000,000 CAD [16]. Director of Personnel Selection, Lieutenant-Colonel Buck Blair further ruminated on the costs of maladaptive behaviour in a November 1955 study. He calculated that the Canadian Army had already lost over 80,000 days of service to men Absent Without Leave (AWOL) that year and would lose an additional 100,000 days before January [10]. A 1954 American report similarly concluded that delinquency cost the armed forces time and money. Co-authors Ira Cisin and Meredith P. Crawford concluded that on average, two per cent of United States soldiers were AWOL at any one time. This meant that nearly 20,000 out of 1,000,000 went AWOL each month [18].

Beyond the financial appeal of pre-emptive screening, the Second World War had also generated unprecedented levels of data. This information could be used to identify more accurate predictors of future behaviour. The introduction of new powerful processing technology from companies like the International Business Machines Corporation (IBM) also meant that it was far easier to record, validate, and compare data than ever before [50].

In other words, it was in the military's interest to keep improving personality assessment. While he remained sharply critical of screening, Franklin Chalke [16] acknowledged that, "continued ignorance of sound personality...screening procedures is wasteful in dollars, detrimental to planning and harmful to the human being involved."

Throughout this period, scientific advancement became pivotal to achieving military superiority, including both the development of new weaponry and the enhancement of human performance. The defence scientific community saw the mind as the key to accessing and perfecting the most fundamental weapons system of them all, man. Militaries around the world formalised and integrated the human sciences as an essential component of research and development [28]; [46]; [60]. In a 2007 article, "Science in Dachau's Shadow," American scholar Alfred McCoy maintains that "effort[s] to crack the code of human consciousness [led to] a veritable Manhattan Project of the mind"([47], p. 402). By the early 1960s, the United States (US) alone invested close to a billion dollars a year in psychological experimentation [47]. Much of this research remained the preserve of PSOs, but the postwar era also witnessed the rise of civilian research agencies established to carry out more rigorous test development and inject a level of intellectual scrutiny to the process. New agencies included Canada's Defence Research Board, founded in 1947 and the American Human Resources Research Office (HumRRO), created several years later in 1951 [20]; [60]; [26].

During the Cold War, personality testing became an essential component of military screening at induction, and during selection for specialised occupations and/or roles (e.g., officers, clearance divers, pilots, special forces etc.) [43]. Driven by Second World War data, the majority of this early research focused on determining how well initial assessments correlated with a serviceman or woman's subsequent performance in training and battle. For example, Canada's Defence Research Board tracked the careers of over 3,000 Canadian service personnel who served between 1939 and 1945. Psychologists used this pool of subjects to better understand personality differences between service members [49]. By the early 1950s, all three nations had also embarked on longitudinal studies of the cohort enlisted since 1945, with the hope of creating selection instruments finely attuned to evolving requirements [6]; [4]; [38]; [57].

Military officials primarily wanted tests that could accurately predict inefficiency, counterproductivity, and even spot outright criminality [36]; [62]. Current integrity testing emphasises the identification and measurement of positive, productive behaviours, or characteristics like conscientiousness [3], [25]. In contrast, early Cold War research focused on delinquency. For instance, the British Army's Operational Research Group completed a series of studies throughout the early 1950s on the military utility of the delinquent. In one such July 1951 study, psychologist Dr. J.C. Penton traced 2,000 individuals released from military prisons and detention barracks in 1945. Penton and his colleagues used the data collected to estimate the cost to the armed forces based on time lost to acts of delinquency, criminality, punishment, and sickness [57]. Across the Atlantic, the United States Provost Marshal commissioned a survey comparing 616 men confined to stockades to a control population of 1,200 soldiers on regular duty. Like their British colleagues, American officials hoped to identify which personality traits most clearly distinguished offenders from the bulk of average recruits [1].

With this in mind, all three nations experimented with a variety of psychometric instruments designed to do just that. Like their First and Second World War predecessors, they favoured the use of questionnaires such as the Cornell Selectee Index and the Minnesota Multiphasic Personality Inventory (MMPI), originally developed to identify and measure abnormal psychopathology [6], [8], [19], [58], [73]. Devices like this helped to quantify complex and shifting theories of personality. In the article "Ink Blots or Profile Plots," Australian historian Roderick Buchanan contends that an inventory "like the MMPI lent [itself] to actuarial-empirical interpretation ([12], p. 187-188). [The] profile patterns could be explicitly defined and justifiably isolated as predictors to be matched with descriptive information. [These] profile patterns could thus be used to generate unambiguous

interpretive hypotheses.” Equally important, the tests could be easily administered to large groups and scored with minimal effort. It is also vital to recognise that the armed forces could purchase test booklets and scoring keys inexpensively because of their widespread use in industry and the private sector [12], [17]. The same cannot be said of projective instruments like the Rorschach Ink Blot or the TAT. Reliant on subjective interpretation, they could not be administered to large groups without complication or understood without specialised training [58].

Experimentation and the introduction of new tests during the 1950s and 1960s yielded results that had a long-lasting impact. During this period, both civilian and military studies laid the foundations of how subsequent generations have understood the average delinquent, in or out of uniform. The typical offender was young, with a history of truancy and patchy employment [16], [35], [55], [57]. Immature and impulsive, they lacked stable parental figures or a sense of family cohesion [4]. In keeping with the precedent set during the First and Second World Wars, researchers argued that the screening process should continue to reject men with a personal or family history of psychosis or major mental illness. They further argued that, when possible, psychometric testing should also be used to winnow out any behaviour prejudicial to good order and discipline, from habitual bedwetting to chronic drunkenness and depression [16]. Guidelines developed for the Canadian Armed Forces in 1954 even counselled PSOs and medical officers to be on the watch for “extremely odd or eccentric attitudes, mannerisms, habits or appearance,” and those men who appeared, “overly meticulous, overly ambitious [or] excessively dependent [16].”

Military studies universally drew direct links between youthful misbehaviour and adult criminality. In April 1951, the Canadian Army’s Committee on Morale Problems highlighted, “the need for unusual effort in trying to make adequate soldiers of present-day youths, particularly those from cities [15].” Similarly, to their Canadian counterparts, the British issued a number of technical reports on juvenile delinquency and its implications for recruitment throughout the early 1950s. For instance, in “The Military Value of the Ex-Borstal Boy [58],” Dr. J.C. Penton argued that a higher percentage of juvenile offenders presented with psychiatric symptoms than a comparable control population (22 per cent versus eight per cent). Writing to the Secretary of State for War in March 1958, Adjutant General Sir Charles Loewen further supported this contention, arguing that “the 5% ‘bad hats’ we seek to eliminate...are responsible for more than half of the cases of crime [44].”

The resulting screening methods and entrance standards owed much to contemporary social mores and traditional paradigms of masculinity. For example, early Cold War studies consistently identified homosexuality as *aberrant*, and characterised *feminine* personality traits as a liability and a drain on combat effectiveness [16]; [63]. Meanwhile, related work on the personality of the *ideal fighter* incessantly valorised stoicism and aggressive masculinity [7]. The Canadian authors of one Korean War era report argued that effective combat soldiers were less emotionally sensitive and anxious than their peers. They enjoyed outdoor pursuits and physically demanding activities that required “considerable expenditure of energy, initiative and occasional daring” and consistently described themselves positively as, “mature, practical...[and] adventurous.” In contrast, non-fighters enjoyed indoor, white-collar pursuits and characterised themselves as, “artistic, dependent, [and] fearful” [66]. The United States military’s Fighter Factor Questionnaire similarly scored interest in outdoor and other *manly* pursuits favourably, and their British counterparts went so far as to devise a three-point scale of aggression to assess the combat temperament of recruits [42]; [58].

Postwar anxieties about a steep rise in rates of juvenile delinquency, crime, and social discord informed which qualities military officials and the defence scientific community valued and shaped how they approached test development. Commentators linked the displacement of the war years to rising levels of separation and divorce and argued that high birth rates, elevated social mobility, and the number of women entering the workforce fuelled family breakdown and youth criminality [33], [78], [74]. Governments on both sides of the Atlantic

subsequently launched numerous enquiries. For instance, the United States Senate began an investigation into juvenile delinquency in 1953 that lasted well over a decade [74]. In the early 1960s, the British Home Office convened a committee on the same subject that led to two major white papers and laid the foundations of the 1969 *Children and Young Persons Act* [78]. The Canadian Department of Justice similarly published a landmark report entitled *Juvenile Delinquency in Canada* in 1965 [27]. More widely, the work of popular clinicians further entrenched the perception that troubled children overwhelmingly become disturbed adults [40], [65]. For example, British psychiatrist John Bowlby's research on maternal attachment directly cited a lack of motherly love in early childhood as the main factor leading to "affectionless criminals; psychopaths; and neurotic, aggressive, oversexed and anxious individuals" ([74], p. 404).

2.4 CHANGING DIRECTION (1970s – 1990s)

The development of personality measures to predict and counter deviant or delinquent behaviour reached its zenith in the 1970s. Rates of drug abuse and indiscipline in all three countries had driven additional investment, particularly in the United States. By 1971, one in every four American soldiers was either AWOL or had deserted [4]. Initially fuelled by American participation in the Vietnam War, the introduction of an all-volunteer force in 1973 further motivated concern for the quality of incoming recruits [61]. In a 1979 article for the *Wilson Quarterly*, celebrated sociologist Charles Moskos noted that the "desertion rate in the all-volunteer force is twice as high as that in the pre-Vietnam era ([51], p. 138)." What is more, the United States Armed Forces experienced record rates of courts martial during the Vietnam War [61].

The United States military eventually discovered that the steep decline in the quality of incoming personnel had largely resulted from the severe mis-norming of the Armed Services Vocational Aptitude Battery (AVSAB) [75]. Revelations surrounding the AVSAB served as a stark reminder that neither intelligence nor personality testing was without flaw. Furthermore, psychometric methods did not always produce reliable, consistent data [61]. Changing recruitment needs throughout the 1970s also prompted the entire defence scientific community to pause and reflect on several decades of research and test development. Reviewing these results, American, Canadian, and British officials finally began to acknowledge the pitfalls of a decades-long enterprise and the need to reconsider the road ahead.

The search for a personality test that could accurately identify deviant or delinquent behaviour remained elusive. Projective instruments proved too difficult to deploy inexpensively. Inventories produced quantifiable data but as in the past, there was significant margin for error. Satisfactory, if not exemplary, candidates could be easily screened out. The "black spot hypothesis" ([1], p. 18) upon which the majority of military screening rested functioned on the premise "that the best estimate of an individual's behaviour in the future is what he has done in the past" ([1], p. 18). While this is arguably true, psychometric measures at that time made little allowances for changes in behaviour [41], [69]. As Bell and his colleagues concluded in a 1978 American review of research on military delinquency ([5], p. 1e):

While the idea that potential troublemakers can be identified before they enter the service is appealing, data do[es] not support the practicality of such an approach. Manipulation of screening standards to reduce military delinquency is possible only at the cost of rejecting large numbers of good men. Therefore the search for predictive individual variables should probably end.

To further complicate matters, the operating characteristics of many selection batteries were woefully out of date by the early 1980s. Military testing procedures primarily reflected data gathered about the generation who had served from 1939 to 1945. Significant demographic, social, cultural, and economic changes had irreversibly altered the composition of the recruiting pool and the norms upon which testing should be based. What is more,

neither military nor civilian psychologists shared a common theory of personality or could even agree to a stable constellation of traits to even measure. This hampered the operationalisation of the resulting data supporting test development. An underlying tension also remained as to whether it was even possible to scientifically measure personality. As one CACDS report [19] released in the 1950s noted, “[was] man’s individuality...too many sided...in its interest, needs, aspirations and character qualities, ever to be reducible to a statistic?”

It is clear that the roots of today’s integrity tests and their use in the military lies squarely in the 20th century, from the First World War to the Cold War [14]. Like other measures of personality, integrity tests “claim to measure an individual’s propensity to behave in certain ways ([71], p. 36).” Moreover, these devices seek to weed out those who would behave in a socially unacceptable manner and identify those who will exhibit positive characteristics like conscientiousness and honesty. However, existing personality testing batteries did not live up to expectations by the late 1970s; over the following decades, there were consistent efforts to seek alternative methods of detection and measurement, including the use of polygraph examination. This type of testing “measures several...physiological indicators of arousal” and any changes “exhibited in response to a set of questions are taken to indicate deception or truthfulness” ([70], p. 11). Used for both criminal investigation and employment screening, employment of polygraph testing increased rapidly throughout the 1970s and 1980s. This included the US Department of Defense, which saw the polygraph as a convenient way to conduct security clearance interviews. But repeated congressional and judicial reviews concluded that “no overall measure or single, simple judgment of polygraph testing validity can be established based on available scientific evidence” ([70], p. 4). There has also been widespread concern that use of such tests for personnel screening could be “susceptible to: 1) countermeasures by persons trained to use physical movement, drugs, or other techniques to avoid detection as deceptive; and 2) false positive errors where innocent persons are incorrectly identified as deceptive” ([70], p. 5). Polygraph tests have also been criticised “for interfering with the rights of private citizens, and for lowering employees’ morale” [70].

Without definitive evidence to support the use of devices like the polygraph, the development of more traditional paper-and-pencil personality and integrity tests continued to evolve. Much of this was driven by demand from the private sector, which sought to use integrity measures in selecting applicants for employment and also wished to lower rates of absenteeism, tardiness, and other counterproductive behaviours like theft [71]. Beginning in the 1980s, the introduction of new models of personality, like the Big Five, also significantly shored up the theoretical foundations of testing [34], [56]. In addition, continual advances in technology made it substantially easier to establish the validity of testing methods and the capacity to carry out large-scale longitudinal studies [34]. The military, like other large institutions, has eagerly adopted these methods. After all, responsive and efficient screening systems remain no less imperative today than in previous decades. Mistakes cost dearly in an operational environment that remains dynamic, competitive, and unforgiving.

But integrity tests and the wider field of personality testing remain contested. While progress has been made, defining and measuring personality, including constructs like honesty, remains a complicated business. Predicting future behaviour and assessing individual potential will always be accompanied by a measure of doubt. Personality testing now rests on more solid theoretical foundations than ever before. Having said that, the process of test development will never be entirely free of subjective judgement and will always be informed by the society in which it takes place. Throughout the twentieth century, socio-cultural standards have played a critical role in how large institutions like the armed forces define personality and select desirable traits with the object of producing “efficient soldiers, sailors and airmen well-adjusted to Service Life [20].” This also includes the rejection or even medicalisation of *undesirable* qualities by creating induction standards that discriminate against those whose behaviour is categorised as socially unacceptable. In addition to this, there are also concerns about privacy or falsely identifying an individual as a “high risk to commit dishonest acts” ([71], p.12) and the stigma that the person may carry as a result.

2.5 CONCLUDING REMARKS

This chapter has examined the history of personality and integrity-related testing in the military. Identifying the 20th century roots of these fields, this piece has also explored how the field has evolved up until the 1990s. The following chapters in this report will engage more directly with developments since that time and look at how approaches to the issue of integrity have since changed, as well as evaluate the current state of scholarship and its application in military settings.

2.6 REFERENCES

- [1] Angle, D.C., and Castelnovo, A. E. (1978). The correlates of AWOL: 30 years of research into the problem of AWOL (Technical Report 78-6). US Army Research Institute for the Behavioural and Social Sciences.
- [2] Ahrenfeldt, R.H. (1958). *Psychiatry in the British Army in the second world war*. Routledge. doi.org/10.4324/9780429445811.
- [3] Ash, P. (1991). Comparison of two integrity tests based upon youthful or adult attitudes and experiences. *Journal of Business and Psychology*, 5(3), 367-381.
- [4] Bauer, R.G., Stout, R.L., and Holtz, R.F. (1976). Developing a conceptual and predictive model of discipline in the US army (Technical Report 76-5). US Army Research Institute for the Behavioural and Social Sciences.
- [5] Bell, D. E. (1975). Summary of ARI research on military delinquency (Research Report 1185). US Army Research Institute for the Behavioural and Social Sciences.
- [6] Blair, W. (1949). A comparative study of offenders and non-offenders in the Canadian Army Masters Thesis, University of Alberta.
- [7] Blair, W. (1955, Feb 21). Characteristics of fighters and non-fighters. Library and Archives of Canada, LAC: R112, Vol. 32173, File No. C-3077-4 Pt. 1.
- [8] Blair, W. (1958, Mar 18). Identification of incipient misconduct problems. Library and Archives of Canada, LAC: R112, Box 822900, File C-2076-2.
- [9] Blair, W. (n.d.). Prediction of military delinquency. Library and Archives of Canada , LAC: R112, Vol. 32174, File C-3077-8, Part 1.
- [10] Blair, W. (1955, Nov 10). Memorandum: AWL and detention. Library and Archives of Canada, LAC: R112, Vol. 32174, File C-3077-8, Part I, HQC 3077-8.
- [11] Bourke, J. (2000). *An intimate history of killing: Face to face killing in twentieth century warfare* (Revised edition). Basic Books.
- [12] Buchanan, R.D. (1997). Ink blots or profile plots: The Rorschach versus the MMPI as the right tool for a science based profession. *Science, Technology, & Human Values*, 22(2), 168-206.

- [13] Butcher, J.N. (2012). Clinical personality assessment: History, evolution, contemporary models, and practical applications. Oxford Handbooks Online.
- [14] Camara, W.J. and Schneider, D.L. (1994). Integrity tests: Facts and unresolved issues. *American Psychologist* 49(2), 112-119.
- [15] Canadian Army (1951, Apr). Final Report of Committee on Morale Problems, 5 April 1951, Library and Archives Canada: R112, Vol. 32174, File C-3077-8, Part 1, HQC 3077-8 (Org 3).
- [16] Chalke, F.C. (n.d.). Preliminary report no. 1: Psychiatric aspects of manpower conservation: Psychiatric screening of recruits. LAC: Box 7564, DRBS 9425-05.
- [17] Chalke, F.C. (1954). Uniform system of manpower classification. Library and Archives of Canada, LAC: RG 24, Acc 1983-1984/167 GAD, Box 7971, Volume I, File 2-6640-2.
- [18] Cisin, I.H., and Crawford, M.P. (1954, Apr). A Preliminary Investigation of Delinquency in the Army. HumRRO Technical Report No. 5 . Defense Technical Informational Center: AD0029029.
- [19] Commonwealth Advisory Committee on Defence Science (1950). The purpose and principles of defence operational research: Memorandum by the United Kingdom delegation. The British National Archives at Kew, TNA: DEFE 10/11.
- [20] Commonwealth Advisory Committee on Defence Science (1950). Memorandum by South African delegation, methods of personnel selection. The British National Archives at Kew, TNA: DEFE 10/11.
- [21] Copp, T. (1996). Selective reasoning in WWII: Army. *Legion Magazine*. <https://legionmagazine.com/en/category/canadian-military-history-in-perspective/army/page/24/>
- [22] Copp, T. (1998). From neurasthenia to post-traumatic stress disorder. In P. Neary, and J. Granatstein (Eds.), *The Veterans' Charter and post-World War II Canada*. McGill Queens University Press.
- [23] Copp, T. and McAndrew, B. (1990). *Battle exhaustion: Soldiers and psychiatrists in the Canadian Army 1939 – 1945*, McGill Queens University Press.
- [24] Copp, T., and Osborne Humphries, M. (2010). *Combat stress in the 20th century: The Commonwealth perspective*. Canadian Defence Academy Press.
- [25] Coyne, I., and Bartram, D. (2002). Assessing the effectiveness of integrity tests: a review. *International Journal of Testing*, 2(1), 15-34 .
- [26] Crawford, M. (1984). Highlights in the development of the Human Resources Research Organization (HumRRO), *American Psychologist*, 39(11), 1267-1271.
- [27] Department of Justice, Canada (1965). *Juvenile delinquency in Canada: The report of the Department of Justice committee on juvenile delinquency*. Department of Justice (Canada).
- [28] Driskell, J. E., and Olmstead, B. (1989). Psychology and the military: Research applications and trends. *American Psychologist*, 44(1), 43-54.

- [29] Egan, J.R., Jackson, L., and Eanes, R.H. (1951). A study of neuropsychiatric rejectees. *Journal of the American Medical Association*, 145(7), 466–469. doi.org/10.1001/jama.1951.02920250018004.
- [30] Fitzpatrick, M. (2017). *Invisible scars: Mental trauma and the Korean War*. University of British Columbia Press.
- [31] Fussell, P. (2013). *The Great War and modern memory*. Oxford University Press.
- [32] Gamble, R.E., and Zickar, M.J. (2008). A history of the early days of personality testing in American industry: An obsession with adjustment. *History of Psychology*, 11(3), 164-184.
- [33] Gleason, M. (1997). Psychology and the construction of the ‘normal’ family in postwar Canada, 1945 – 1960. *Canadian Historical Review*, 78(3), 442-477.
- [34] Goodstein, L., and Lanyon, R.I. (1999). Applications of personality assessment to the workplace: A review. *Journal of Business and Psychology*, 13(3),291-322.
- [35] Haldane, I. (1951, October). Psychological testing of delinquents (AORG Report No. 18/51). TNA: WO 291/1150.
- [36] Hall, W., and Eaves, J. (1990). *Military psychology: Its role in the growth of psychology in Australia*. 1st Australian Psychological Research Unit.
- [37] Howard, J. (n.d.). Some notes on the selection interview. Library and Archives of Canada, LAC: R10916, Volume 1, File 13.
- [38] Hymes, J.P., and Blackman, S. (1965). Situational variables in socially deviant behaviour. *The Journal of Social Psychology*, 65(1), 149-153.
- [39] Jones, E., and Wessely, S. (2005). *Shell shock to PTSD: Military psychiatry from 1900 to the Gulf War*. Taylor & Francis Group/Psychology Press.
- [40] Kafka, J. (2008). Disciplining youth, disciplining women: Motherhood, delinquency and race in postwar. *American Educational Studies Association*, 44(3), 197-221.
- [41] Larson, E.E., and Kristiansen, D.M. (April 1969). Prediction of disciplinary offense early in military service (Technical Research Note 210). US Army Behavioral Science Research Laboratory.
- [42] Major A.B. Lavar, Personnel Selection Service to Major A.R. Pichem Infantry Requirements Project, 7 February 1962, LAC: R112, Vol. 32173, File C-3077-5, Part I.
- [43] Leese, P. (2002). *Shell shock: Traumatic neurosis and the British Soldiers of the First World War*. Palgrave Macmillan.
- [44] Adjutant General Sir Charles Loewen to Secretary of State for War, 3 March 1958, TNA: WO 32/16185.
- [45] Lunstrum, E. (2012). The politics of psychology in the British empire, 1898-1960. *Past & Present*, 215(1), 195-233.

- [46] Maier, M.H. (1993). Military aptitude testing: The past fifty years (Technical Report 93-007). Defense Manpower Data Center.
- [47] McCoy, A.W. (2007). Science in Dachau's Shadow: Hebb, Beecher, and the development of CIA psychological torture and modern medical ethics. *Journal of the History of Behavioral Sciences*, 43(4), 401-417.
- [48] Menninger, W.C. (1948). *Psychiatry in a troubled world, yesterday's war and today's challenge*. Macmillan.
- [49] *Military Psychology* 3. (1953, Apr 13). LAC: DRBS 2-150-302.
- [50] Mooney, C. (1953, Jun 18). Defence Research Medical Laboratories to Chief of the Air Staff, Attention: Group Captain P.A. Gilchrist, WD Release Rate. LAC: DRBS 68-153-301.
- [51] Moskos, C. (Spring 1979). The all-volunteer force. *Wilson Quarterly* 3.2.
- [52] Myers, D. (2012). *Social psychology* (11th edition). McGraw-Hill Education.
- [53] National Army Museum. (2021). World War 1. <https://www.nam.ac.uk/microsites/ww1/>
- [54] Natzio, G. (1993). British Army servicemen and women 1939 – 1945: Their selection, care and management. *Royal United Services Institute Journal*, 138(1), unknown pages.
- [55] Nichols, R.S. (1961 – 1962). Progress report: First court martial screening program pilot study. US Mental Hygiene Consultation Service.
- [56] Oh, I.S., Lee, K., Ashton, M.C., and de Vries, R.E. (2011). Are dishonest extroverts more harmful than dishonest introverts? The interaction effects of honesty-humility and extraversion in predicting workplace deviance. *Applied Psychology*, 60(3), 496-516.
- [57] Penton, J.C. (1951, Jul). A study of the subsequent careers from the british liberation army released from military prisons and detention barracks under suspended sentence (Report 9/51). Army Operational Research Group. TNA: WO 291/1141.
- [58] Penton, J.C. (1951, Nov). Military value of the ex-borstal boy (Report 9/51). The British National Archives at Kew: WO 291/1152.
- [59] Penton, J., and Haldane, I. (1955, Nov). A study of the evidence upon which the psychiatric element of other ranks is based (Memorandum No. 20). British Army Operational Research Group. TNA: WO 291/1401.
- [60] Prociuk, T.J. (1988). Applied psychology in the Canadian Forces: An overview of current research. *Canadian Psychology*, 29(1), 94-102.
- [61] Rimland, B., and Larson, G.E. (1981). The manpower quality decline: An ecological perspective. *Armed Forces & Society*, 8(1), 21-78.
- [62] Sackett, P.R. (1994). Integrity testing for personnel selection. *Current Directions in Psychological Science*, 3, 73-76.

- [63] Sarbin, T.R. (1991). Homosexuality and personnel security. Defense Personnel Security Research and Education Center.
- [64] Scroggins, W.A., Thomas, S.L., and Morris, J.A. (2008). Psychological testing in personnel selection, Part I: A century of psychological testing. *Public Personnel Management*, 37(1), 99-109.
- [65] Scull, A. (2015). Contending professions: Sciences of the brain and the mind in the United States, 1850 – 2013. *Science in Context*, 28(1), 131-161.
- [66] Services selection report: Instructions for completion (n.d.). Library and Archives of Canada, LAC: R112, Vol. 32174, File 3077-7 Pt. I.
- [67] Shephard, B. (2001). *A war of nerves: Soldiers and psychiatrists in the twentieth century* (1st edition). Harvard University Press.
- [68] Skomorosky, A. (2013). Occupational personality: Implications for military organizational settings. In P. Johnston, and K. Farley, *Military human resources issues: A multinational view*, 71-83. Winnipeg, MB: Canadian Defence Academy Press.
- [69] Uhlaner, J. (1976). Predicting military delinquency (Research Problem Review 76-4). US Army Research Institute for the Behavioural and Social Sciences.
- [70] United States Congress Office of Technology Assessment (1983). *Scientific Validity of Polygraph Testing: A Research Review and Evaluation – A Technical Memorandum (OTA-TM-H-15)*. <https://www.ota.fas.org/reports/8320.pdf>
- [71] United States Congress Office of Technology Assessment (1990). *The use of integrity tests for pre-employment screening (OTA-SET-442)*. <https://www.ota.fas.org/reports/9042.pdf>
- [72] Vearing, A., Wheeler, J., and Kehoe, E. J. (2013). Finding the right people: Taking the guesswork out of selection. In P. Johnston, and K. Farley (Eds.), *Military Human Resources Issues: A Multinational View*, 49-60. Canadian Defence Academy Press.
- [73] Vernon, P. (1947). Research on personnel selection in the Royal Navy and the British Army. *American Psychologist*, 2(2), 35-51.
- [74] Vicedo, M. (2011). The social nature of the mother's tie to her child: John Bowlby's theory of attachment in post-war America. *British Journal for the History of Science*, 44(162, Part 3), 401-426.
- [75] Walkawicz, J. S. (2009). Tapping into the US cultural shift: revising enlistment standards today to avoid conscription tomorrow [Master's Thesis]. US Joint Forces Staff College.
- [76] Witmer, H., Arend, I., Edwards, M., and Kranz, J. (1943). A plan for helping to identify men psychologically unadapted to military service. *Smith College Studies in Social Work*, 13(4), 299-336.
- [77] Woodworth, R. (1921). Psychological test, personal data sheet, Woodworth. National Museum of American History.

- [78] Wright, M. (1970). Twenty years of the British journal of delinquency/criminology. *British Journal of Criminology*, 10(4), 372-382.

Chapter 3 – CURRENT STATE OF MILITARY PRACTICE AND RESEARCH ON INTEGRITY ASSESSMENT

Marjolein Verboom
Netherlands Armed Forces
NETHERLANDS

Kai Nyman
Finnish Defence Research Agency
FINLAND

Maria Fors Brandebo
Swedish Defence University
SWEDEN

Françoise Bertrand
Directorate General Human Resources
Belgian Defence
BELGIUM

3.1 INTRODUCTION

In this chapter, we will examine if and how integrity testing is incorporated in the military selection processes of 11 participating countries of this RTG (Australia, Belgium, Canada, Czech Republic, Denmark, Finland, United Kingdom [British Army], Germany, the Netherlands, New Zealand, and Sweden). We will then provide a closer analysis of the research being currently conducted in a number of these nations towards the development of integrity assessment.

3.2 CURRENT PRACTICE

A Request for Information (RFI) was used to collect information on the current methods of integrity assessment for military selection. Members of the Integrity Group of the RTG composed this RFI and dispatched it to all thirteen participating countries. An overview of the questions posed in the RFI is provided in Table 3-1. A total of 11 countries responded, which are accordingly listed in Table 3-2.

The questionnaire consisted of seven questions structured to collect as much relevant information as possible. To get a fulsome overview of the current situation, the RFI included a range of questions on integrity assessment in military selection. This included questions related to current methods, scoring systems, challenges encountered, and lessons learned. This section describes the responses to questions in Table 3-1, gathered through this RFI.

Table 3-1: RFI Questions on Integrity Testing.

Does a nation currently assess integrity (directly) or through related indicators/predictors (e.g., personality traits, attitudes, values, past behaviours)
If not, explain why your nation does not assess integrity?
What constructs, personality dimensions, or instruments/scales are being used to assess integrity?
What selection methods are being used to assess integrity?
How is information from integrity-related selection measures/methods used to make selection decisions (e.g., structured scoring system, expert judgement)?
Difficulties or challenges in assessing integrity?
Lessons learned with respect to assessing integrity

3.2.1 Methods of Integrity Assessment

As might be expected, all participating countries currently include some measure of integrity (or integrity-related factors) as part of their core selection process. All countries rely upon indirect (I) methods to gain an impression of an applicants’ integrity rather than using direct measures. Although Belgium has developed a psychological test / instrument that claims to directly measure integrity, they are not presently using it in selection. At this juncture, it is important to note that while background security screening may provide insight into the integrity of a candidate, this chapter focuses on the assessment of integrity within selection. Table 3-2 illustrates how integrity is included as a measurement in the selection process of the participating nations.

Table 3-2: Overview Measuring Integrity During Selection.

Country	General-Entry Level	Special Occupations
Australia	I	I
Belgium	I	*
Canada	-	I
Czech Republic	I	I
Denmark	-	I
Finland	I	I
Germany	I	I
United Kingdom [^]	I**	I
The Netherlands	I	I*
New Zealand	I	I
Sweden	I	I

I = indirect methods; ^ = British Army only.

*** All candidates for special occupations take the general-entry level test before.**

**** This pertains to the Army only; the Air Force doesn’t measure integrity, and information on the Navy is missing.**

If an applicants’ integrity is not assessed directly through a purpose-built test, how is it being judged? It appears that in most European nations, information relating to integrity is generally collected via a semi-structured interview with a psychologist or trained assessor, to assess suitability for military service. In Canada, the security screening process is generally assumed to provide insights into an applicants’ integrity. The central paradigm underpinning each country’s interview model is that ‘the best predictor of future behaviour is past behaviour’. In other words, the assessing psychologist is looking for evidence that applicants have behaved in a manner consistent with that country’s values in the past, including integrity, or have demonstrated the capacity to do so in the future.

3.2.2 Contribution to the Selection Decision

All countries currently use indirect methods (e.g., suitability interview) to make inferences about integrity. As such, this section summarises the approaches used by these armed forces in compiling information to inform selection

decisions about suitability or potential for service. To reach a final decision about an applicant’s suitability for military service, the psychologist considers multiple sources of information generated from the assessment and uses their expert professional judgement to make a final recommendation. To make an objective assessment, the psychologist consults the results of all tests and questionnaires administered and compares those results to other information provided in the interview. Inconsistencies may mean that a second interview is required.

Due to diverse views on the integrity construct, assessment and scoring processes currently entail varied approaches. Most countries examine an applicant’s past behaviour, their stated attitude towards that behaviour, and consider this as an indicator of their likely integrity in the future. Other Countries like Belgium, Czech Republic, Finland, Germany, and Sweden use cut off scores on various measures (e.g., self-report personality measures, situational judgement tests, biographical questionnaire, job simulation) to ascertain suitability for enlistment or employment. Canada currently relies on the background security screening process to provide an indication of applicants’ integrity; however, scores from two personality factors (Conscientiousness, and Emotional Stability), which appear to be used as indicators of integrity in other armed forces, are combined with scores from other measures to determine military potential in general-entry selection. Table 3-3 shows how the final decision on suitability for employment, based on integrity, is made in the participating countries.

Table 3-3: Overview of Methods Used to Assess Integrity Information in the Selection Decision.

Country	Standardised Scoring System	Expert Judgement	Use Cut off Score on Test Results
Australia	x	x	
Belgium	x	x	
Canada		x (security screening)	
Czech Republic		x	x
Denmark			
Finland	x	x	x
Germany	x	x	x
United Kingdom^	x	x	
The Netherlands	x	x	
New Zealand			
Sweden	x	x	x

^British Army only

Guidelines to assist the psychologist to make their decision are usually included in an operating or technical manual. These guidelines help standardise both the quality of the assessments conducted, and the decisions made as a result of those assessments. In some countries like Australia, Belgium, Germany, the Netherlands, Sweden and New Zealand, the standardisation process includes a structured rating or scoring process to help minimise potential subjectivity in the final decision.

In the case of officer selection or specific occupation selection, a secondary selection process diverging from the basic selection process is often employed. This 'selection board' usually involves a group of assessors, including a psychologist. Group assessment processes are generally governed by similar operating guidelines that help standardise the process and the decisions arising from it.

3.2.3 Difficulties and Challenges with Integrity Assessment

Maybe integrity is something like a butterfly, we cannot catch it, we can just observe it, watch it. It is something 'fragile', so we are not able to measure it straightly or properly, we can only think about it, concern it, contemplate about it.

Markéta Zajícová, Czech Republic

As discussed in a later chapter, integrity can be difficult to clearly define. The broad range of constructs that underlie this concept, as operationalised by participating countries, remains one of the biggest problems faced in assessing integrity. The most frequently reported difficulty in measuring integrity is definitional clarity or the lack of an agreed, shared understanding of what integrity constitutes. This lack of a single agreed definition means that the majority of military selection systems employ an indirect rather than direct measurement of integrity. This lack of agreement regarding construct definition and clarity is also likely to contribute to the lack of consistency regarding the best approach to follow (e.g., personality tests, semi-structured interview) in assessing integrity.

The participating countries have also identified several other key challenges in integrity measurement. These include challenges with scoring existing measures of integrity (i.e., compiling information from varied sources in a reliable and valid manner), issues with impression management and social desirability of some measures, difficulty obtaining psychologists with the requisite education and training to conduct interview assessments, lack of resources to conduct necessary research to develop suitable measures of integrity, resistance amongst military practitioners to include additional measures of integrity in what is already a rigorous selection or 'weeding-out' process, unless significant material gains can first be demonstrated.

Cost-benefit considerations must also be considered in the development of a targeted and objective integrity test and questions must be asked as to whether such a measure would add sufficient value to the existing semi-structured interviews conducted by trained professionals. The investment in research and development required to develop and maintain such an instrument would be significant. Other additional cost factors might include costs associated with publication and supply of test materials; psychologist and/or test administrator training and accreditation; and candidate and psychologist time and effort. These factors would need to be considered when evaluating the costs versus the benefits of adding a new test to existing selection methods.

Interpretation of performance on an integrity test might also prove difficult. For example, if a candidate answers questions truthfully but in doing so discloses past breaches and transgressions, what does it mean for the person's integrity? Should that person be excluded for being truthful on the test, when someone who is less truthful would not be similarly punished? If test items are too transparent, this dilemma is compounded by the likelihood of applicants responding to the test in a socially desirable way.

3.3 ONGOING INTEGRITY RESEARCH

Four nations (Belgium, Canada, Finland, and Sweden) report that they are currently conducting research related to integrity. This section summarises the research undertaken by these countries.

3.3.1 Belgium

In their work on improving selection processes for Belgian Defence, Bertrand and colleagues [1] developed a Situational Judgement Test (SJT). The first step in this process was to identify relevant competencies and associated behavioural indicators. To conduct a job analysis, the authors employed two approaches that can be characterised as: 1) bottom-up, and 2) top-down. In the bottom-up approach, they selected 347 trainees. Using a critical incident method, they then identified various problematic situations that the trainees had encountered in the execution of their work or training. These incidents (a total of 2082) were categorised and linked to one or more competencies identified by a team of psychologists from Belgian Defence. They found 5994 competency classifications. As a last step, they ranked the competencies based on importance and the eight most frequently mentioned were selected for analysis using a top-down approach. At this point, experts with extensive experience evaluating trainees evaluated the importance of each competency.

The bottom-up approach demonstrated that 0.3 to 0.8 % of the critical incidents identified by other ranks, and 0.8 to 1.6 % identified by non-commissioned members and officers relate to the competency of Integrity. In total, five distinct competencies emerged as the most important. Integrity and loyalty did not feature amongst them. Nonetheless, both were included in the top eight, as they were considered key organisational values. Belgian Defence also considers respecting others as a central competency, which covers many of the same indicators as integrity.

Meanwhile, the experts who participated in the top-down phase of this research identified four competencies as more important than the others: Loyalty, Integrity, Flexibility, and Collaboration.

The next step was to construct a Situational Judgement Test (SJT). The aim of an SJT is to assess an applicant's judgement regarding a situation encountered in the workplace. The critical incidents collected in the bottom-up approach were used as a foundation to construct descriptions of situations. Different response options were constructed, and a scoring key developed. The scoring key was created using a mixed method approach. First, five subject matter experts (military instructors) responded to the questionnaire. Afterward, they were asked as a group, to discuss the effectiveness of each answer. Next, they applied the empirical method in which 80 military trainees completed the questionnaire. The trainees were at the same time evaluated by their instructors on the four competencies: Flexibility, Loyalty, Integrity and Collaboration. Based on these scores, the trainees were divided into two different groups, high and low performers. The scores on the response options were then compared between the two different groups. Finally, the scores given by the trainees were compared with those given by the experts. If the scores of the high performers corresponded to the scores tabulated by the experts, the response options were kept, if not, they were adapted or deleted.

The next step of the validation process was to test the SJT with 49 situations for 1270 applicants, and 37 situations for 970 applicants. After each pre-test, the test was revised by removing situations, adapting content, and adapting the scores. In 2018, this revision had resulted in 30 remaining situations.

3.3.2 Canada

The Canadian Armed Forces has pursued several lines of research to identify a suitable measure of integrity for use in selection. The first line of research focussed on identifying personality facets through a construct-oriented literature review. For example, Darr [3] and Darr and Klammer [4] applied Hogan's elements of moral character to rationally infer possible bright personality facets (from within the Five Factor Model of personality), as well as maladaptive or dark personality facets (from within the Personality Inventory of the Diagnostic and Statistical Manual of Mental Disorders [DSM]; see Ref. [10]).

This research led to the identification of 15 personality facets with theoretical relevance to moral character. Darr [5] subsequently empirically examined these 15 facets to identify the strongest ones. She also compared the bright and dark facets based on their relative importance and incremental contribution predicting integrity outcomes or self-reported counter-productive work behaviours. Summarising findings from two different samples (military police and university students), Darr concluded that some facets had higher relative importance across various criteria. Dependability, achievement, and cautiousness, which fall under the Conscientiousness factor, were amongst the strongest bright facet predictors. Irresponsibility (a Disinhibition factor), deceitfulness and callousness (both Antagonism facets) were amongst the dark facets. In determining the relative contribution of dark facets compared to bright ones, her findings suggested that the dark facets had about the same or slightly higher relative importance and contributed incrementally to the prediction of criteria over and above bright counterparts. Hierarchical regression findings, however, also suggested that one to two dark facets can account for as much variance as multiple bright facets.

The second line of research has focussed on the identification of a composite measure of integrity based on items from the Trait Self Descriptive (TSD) personality inventory. This is a measure of the Five Factor Model (FFM) of personality that is currently used in CAF selection. In brief, Catano, O’Keefe, Francis, and Owens [2] found that a 10-item TSD-based composite, incorporating the personality factors of conscientiousness (4), agreeableness (4), and emotional stability (2), was moderately correlated ($r = -.40, p < .01$) with a self-report measure of counter-productive work behaviour in the expected direction. They also concluded that those scoring high on this composite were more likely to respond in an honest (rather than socially desirable) manner when completing a personality measure. Darr [5] compared the predictive power of the TSD-based integrity composite against the aforementioned important bright/dark facets and a situational judgement test of integrity. She noted that the TSD-based composite had the same or lower importance compared to the other predictor measures against which it was compared. It also accounted for a significant amount of variance in both samples (different criteria) when examined as a predictor by itself; however, significant incremental variance was explained by irresponsibility (both samples) and cautiousness (applicant sample). This suggests that the TSD-based composite may be supplemented by at least two additional predictors, cautiousness and irresponsibility, to fully account for all possible explained variance in relevant outcomes.

Recently, there have been efforts to open up a possible third line of research, which resulted from Darr’s [5] review of research that supports the CAF’s Defence Ethics Program. Her review provided insight into the complex process of moral or ethical decision-making and underscored the importance of both individual and situational factors in fully accounting for variance in counter-productive outcomes. Mainstream research has also highlighted the influential role situational factors play in improving the criterion-related validity of a selection assessment (e.g., Ref. [9]). Selection research is based on the notion of behavioural consistency, and the focus has been on generalisation across individuals and situations [7], [15]. In other words, situational factors have been regarded as moderators or unwanted sources of variance [7]. However, Darr’s [5] review suggests that it may be important to deliberately incorporate relevant situational factors (via the development of tools like situational interviews, situational judgement tests, role play exercises) to improve validity in predicting integrity-relevant outcomes.

3.3.3 Finland

The usefulness of integrity scales has been investigated throughout the development of the Finnish armed forces’ selection system for conscripts. Two different kinds of integrity tests were selected for further examination: i) VIA-integrity scale (IPIP) [12], [13], an overt-type of inventory and ii) TSD-integrity scale [2], a covert-type of inventory. A new test was also developed using both overt and covert subscales. The relationships between integrity and personality traits were studied. The integrity scales were measured in the beginning of military

service and were then compared to military performance ratings at a later time (timeframe between measurements varied from 6 to 12 months). Integrity measures were also compared to officers' ratings of their trainees' leadership potential assessed in the second month of the service. Studies were based on the data of Finnish conscripts ($n = 3441$), who took part to the research voluntarily.

The new integrity scale was a self-report measure based on the military definition of integrity developed by this RTG. The scale was named Military Integrity Model (MIM)-integrity scale. It had six elements (or subscales). These included: Acting in accordance with rules, regulations, and society values; Respecting others; Service before self; Courage; Honesty; Accountability. Examples of items within the subscales included "Act as if there were no rules" and "Do things by the book" (Acting in accordance with rules, regulations, and society values), "Behave properly" and "Trust others." (Respecting others), "Do my tasks only just before they need to be done" and "Do more than what's expected of me" (Service before self), "Don't hesitate to express an unpopular opinion" and "Avoid dealing with awkward situations" (Courage), "Cheat to get ahead" and "Misrepresent the facts" (Honesty), "Keep my promises" and "Dislike taking responsibility for making decisions" (Accountability).

The MIM-integrity scale correlated strongly to the VIA- and TSD-integrity scales. The strongest correlations were between the MIM- and TSD-integrity ($r = .87$) and MIM- and VIA-integrity ($r = .84$). The correlation between VIA- and TSD-integrity scales was $r = .72$. The content of the MIM-integrity scale appears to be so comprehensive that it measures largely the same constructs that the other two integrity measures. Based on previous research, there was an expectation that the Big Five factors of Agreeableness, Conscientiousness and Emotional stability would correlate more strongly with integrity scales than the Big Five scales for Extraversion and Openness. In reality, that expectation was only fulfilled with the VIA-Integrity scale. The MIM-Integrity scale had high correlations with all Big Five scales and the TSD-Integrity scale was most strongly correlated with Conscientiousness, Emotional Stability and Openness. It is assumed that the reason for unexpected high correlations of Openness depends on the content of the Openness scale, which was modified from the original version to better fit to the military context. It looks like the Military Openness scale is more relevant to a person of integrity than the traditional Openness scale is. Relationships between integrity scales and performance ratings were modest in the total group. Results varied widely in different subgroups. In some military tasks, there was a strong association between integrity and performance, while the association was weak in others.

Of the three studied integrity scales, the new MIM scale had the highest correlations in most cases. The military integrity elements "Respecting others" and "Rule following" had strongest associations with the observed military performance from all of the measured elements. The MIM-integrity scale was among the best five predictors, when relationships between performance criteria and 28 psychological constructs were analysed. The other top five constructs were Sense of Coherence, Locus of Control, Dutifulness and Self-Efficacy. The result of the incremental validity analyses was that integrity scales and the Big Five personality measures equally predicted general performance. In the total group, correlations between integrity scales and general performance were modest and correlations between self-rated integrity and observed leader potential were good.

This study concluded that the results support adding an integrity scale into the assessment system of Finnish conscripts. Integrity scales had predictive validity, which varied between subgroups. The predictive validity in the total group was at the same level as Van Iddekinge et al. [17] presented in their meta-analyses, but not as high as in Ones et al. [11] and Schmidt and Hunter [14].

3.3.4 Sweden

Fors Brandebo and colleagues [6] carried out a planned series of studies with the aim to validate the officers' programs selection process. The authors initially aimed to identify qualities and abilities that are of importance

for military officers' ability to perform their duty, regardless of the arena in which they work. The developed a comprehensive model based on:

- 1) A literature study of international peer-review publications;
- 2) A review of Swedish reports and documents with the aim to identify/categorise qualities and abilities in military officers; and
- 3) Focus-group interviews with 39 participants who were deemed to have relevant insight into which qualities and abilities a military officer should possess.

The comprehensive model showed that a military officer should possess a few basic individual qualities such as emotional stability, motivation, determination, energy, integrity, and values. Important interpersonal and task-related abilities were also identified. Integrity was found to be an important quality in military officers in all of the three approaches (literature review, review of reports and documents, interviews).

The second step aimed to further investigate if the identified abilities and qualities from the first study were stable or continue to develop over time, either through education or professional experience [16]. A qualitative study was conducted to assess how these abilities and qualities were perceived to develop. Interviews were conducted with 20 military officers. In this instance, integrity is described as taking responsibility/being loyal, respecting others, and acting as a role model. The first two aspects of integrity were considered to be relatively stable over time, while being a role model was deemed to be subject to development. Most of the participants connected development as a role model to training and work experiences, as well as social interactions with others. The two first aspects, although overall considered stable, was also regarded as developable if the individual has a stable foundation from which to start. The third stage of this project is ongoing and involves a longitudinal study with the goal of investigating if the criteria used to select individuals to the officers' program can predict outcomes like individual performance, motivation to complete the education, and overall health. Integrity is not used as an outcome measure but is measured as a part of the selection criteria (psychologist interview). However, it is unclear at this time if ratings of integrity that can be used in the study (or if the only related source is a weighted score of several criteria).

3.4 CONCLUSION

This chapter has demonstrated that various nations take distinctly different approaches to the assessment of integrity, based on their own notion of what that concept means. The further development of direct measures of integrity will be challenging due to understandable budgetary limits and lack of appropriately qualified personnel. Indeed, Belgium is the only nation out of the eleven countries that has a dedicated test designed specifically for the measurement of integrity. Their Situational Judgement Test uses a series of dilemmas to measure two different aspects of the integrity construct – does the person want to be honest (a *value* variable) and is he/she capable of being honest (a *competence* variable). While the predictive validity of this measure is good, the tool is used only to identify extreme responses, because of issues with transparency and social desirability.

Most countries exploring the possibility of either developing or purchasing an instrument assessing integrity have found it difficult to operationalise this construct in a consistent, meaningful, and practical way. As a consequence, the identification or development of such a measure will remain difficult. The layered or tiered approach to integrity assessment currently employed by most participating nations, utilising different combinations of interviews, personality tests, and potentially other instruments (e.g., an honesty test) might remain the best solution.

Most of ongoing research is related to developing new instruments to predict integrity (or moral character). In this sense, Belgium appears to be ahead in its development of an SJT that assesses integrity, and Finland has operationalised the definition of integrity developed by this RTG to develop its own integrity scale. Sweden, on the other hand, proposes to examine whether the current assessment of integrity (via psychologist interview) predicts later performance and other outcomes.

3.5 REFERENCES

- [1] Bertrand, F., Defranc, A., Huybens, W., De Nil, V., Van Landeghem, K., Tibax, V., Peeters, H., and Mylle, J. (2017). Improving selection: Advances in the Belgian Defence Force. In S.V. Bowles, and P.T. Bartone (Eds.), *Handbook of Military Psychology: Clinical and Organizational Practice*, 261-276. Springerlink.
- [2] Catano, V.M., O’Keefe, D.F., Francis, R.E., and Owens, S.M. (2018). Construct-based approach to developing a short, personality-based measure of integrity. *International Journal of Selection and Assessment*, 26, 75-92. doi: 10.1111/ijsa.12197.
- [3] Darr, W. (2015). Developing a new personality measure: Facet identification (Director General Military Personnel Research and Analysis Scientific Report DRDC-RDDC-2015-R177). Defence Research and Development Canada.
- [4] Darr, W. and Klammer, J. (2016). Personality and destructive leadership: Potential for selection. In D. Watola, and D. Woycheshin (Eds.), *Negative leadership: International perspectives*, 87-118. Canadian Defence Academy Press.
- [5] Darr, W. (2019). Identifying integrity-relevant predictors: An empirical examination of personality facets. Director General Military Personnel Research and Analysis Scientific Report DRDC-RDDC-2019-R154. Defence Research and Development Canada.
- [6] Fors Brandebo, M., Lundell, E., Wolgers, G., and Larsson, G. (2017). Validering av officersprogrammets utvalprocess. Steg 1: Identifiering av individuella förmågor och egenskaper som kan ligga till grund för prediktion av framtida funktionsförmåga [Validation of the officers’ programs selection process. Step 1: Identifying individual abilities and qualities that can predict future performance] (ISSL – Leadership center Series I:109. Karlstad: Swedish Defence University).
- [7] Guion, R.M. (1991). Personnel selection, assessment, and placement. In M.D. Dunnette and L.M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 2). Consulting Psychologists Press.
- [8] Hogan, R. (1973). Moral conduct and moral character: A psychological perspective. *Psychological Bulletin*, 79(4), 217-232.
- [9] Jansen, A., Melchers, K.G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., and König, C.J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, 98(2), 326. doi: 10.1037/a0031257.
- [10] Krueger, R.F., Derringer, J., Markon, K.E., Watson, D., and Skodol, A.E. (2012). Initial construction of a maladaptive personality trait model and inventory for the DSM-5. *Psychological Medicine*, 42(9), 1879-1890. doi:10.1017/S0033291711002674.

- [11] Ones, D.S., Viswesvaran, C., and Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679-703.
- [12] Park, N., Peterson, C., and Seligman, M.E.P. (2004). Strengths of character and well-being. *Journal of Social and Clinical Psychology*, 23, 603-619.
- [13] Peterson, C., Stephens, J.P., Park, N., and Seligman, M.E.P. (2009). Strengths of Character and Work. In N. Garcea, S. Harrington, and A.P. Linley (Eds.), *Oxford Handbook of Positive Psychology and Work*. Oxford University Press.
- [14] Schmidt, F.L., and Hunter, J.E. (1998). The validity and utility of selection methods in personnel selection: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- [15] Schmitt, N., and Ostroff, C. (1986). Operationalizing the “behavioral consistency” approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, 39(1), 91-108.
- [16] Svensén, S., Berglund, A.K., Thielen, A., Wolgers, G., and Fors Brandebo, M. (2018). Validering av officersprogrammets utvalsprocess. Steg 2: En kvalitativ studie om utveckling av egenskaper och förmågor hos officerare [Validation of the officers’ programs selection process. Step 2: A qualitative study on the development of officer’s abilities and qualities]. (ISSL – Leadership center Series I: 114. Karlstad: Swedish Defence University).
- [17] Van Iddekinge, C.H., Roth, P.L., Raymark, P.H., and Odle-Dusseau, H.N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499-530.

Chapter 4 – MEASURING INTEGRITY

Wendy Darr

Department of National Defence
CANADA

Tom H. Skoglund

Norwegian Armed Forces
NORWAY

Markéta Zajícová

Czech Armed Forces
CZECH REPUBLIC

Gerhard Wolgers

Swedish Armed Forces
SWEDEN

This chapter has three main goals: to provide an overview of existing approaches used to assess integrity, to review currently available commercial integrity assessments, and to provide insight into emerging approaches.

4.1 EXISTING APPROACHES TO ASSESS INTEGRITY

In light of the numerous published reviews (e.g., Refs. [74], [75], [98]) on past developments in integrity testing, this section provides only a brief overview on this topic.

4.1.1 Overt and Covert Measures

In general, integrity tests can be categorised as being either overt or covert in nature. The specific content of such tests varies, depending on the definition of integrity and the theoretical background of the test developer. Overt integrity tests, also referred to as “clear purpose” tests, refer directly to dishonest and counterproductive behaviours such as theft, cyber-loafing, and attitudes toward such behaviours [75]. As a result, individuals taking these tests will be well aware of the aim of such measures. Covert tests, also referred to as “disguised purpose” [75] measures, are mostly personality-based. They assess integrity via personality measures like conscientiousness and dependability [98]. Both types of tests have been found to predict overall job performance and counterproductive work behaviour (CWB; [62], [63], [97]). In an updated meta-analysis, the corrected validities of these types of tests in predicting job performance were .14 (overt) and .18 (personality-based), and in predicting CWB were .38 (overt) and .27 (personality-based; [97]). Social desirability or faking, which is the tendency of test takers to distort their responses to manage impressions when completing such measures, accompanies both types of measures, but does not affect the validity of integrity tests [62].

4.1.2 Other Measures

The polygraph and the Minnesota Multiphasic Personality Inventory (MMPI) have also been used to assess integrity but are not typically discussed in mainstream personnel selection literature.

4.1.2.1 Polygraph Testing

The polygraph, or “lie detector” test, involves the use of a device that records several physiological indicators which are used to make inferences about a person’s credibility [75]. A critical review of this test, prompted by the United States (US) Congress in the 1980s, revealed unsatisfactory scientific evidence to support its use in employment testing, and most North American employers were banned from using this type of test in pre-employment testing following the introduction of the Employee Polygraph Protection Act [60], [75]. However, provisions in the Act allow pre-employment polygraph testing in some sectors (e.g., policing, nuclear

and pharmaceutical industries) where public safety is a concern [42], [51]. Consequently, polygraph testing continues to be used in pre-employment screening in the US. In addition, there are reports that countries like Japan, Canada, and Bulgaria also use polygraph testing in pre-employment screening [51], [58].

Results from a comprehensive review of polygraph testing procedures found mixed evidence to support its use in pre-employment screening [58]. However, research into polygraph testing techniques, scoring, and interpretation of results has grown over the years in an effort to develop a set of best practices around its use. Handler et al.'s [42] review on the topic suggests that data-driven decision-making approaches in polygraph testing, use of valid screening selection questions (i.e., those linked to job performance and risk), and a consideration of polygraph results in conjunction with other assessment tools can improve validity, sensitivity, and utility in honesty testing. They acknowledge, however, that polygraph testing is an imperfect tool, and ethical use requires a thorough understanding of its complexities and limitations. The use of polygraph testing in pre-employment screening also requires a consideration of relevant legislation (e.g., human rights, disabilities) that prohibits the collection of certain information (e.g., medical history) prior to a conditional job offer.

4.1.2.2 MMPI

The MMPI is a self-report measure of personality and psychopathology that involves the completion of a questionnaire whose results are interpreted by a clinical psychologist [22]. The MMPI has been used to assess integrity in police officer selection in the US for decades [78]. The validity of the original version of the MMPI in predicting integrity violations in police officer samples was poor [15], but subsequent revised versions (i.e., MMPI-2, MMPI-2-RF) have yielded more promising results. Certain MMPI sub-scales have been found to identify those with an increased risk of problematic behaviours in US police samples (e.g., Refs. [70], [92], [93]).

The MMPI and its variants have also been translated and used in clinical (e.g., patient psychopathology), research (e.g., psychometric examinations), and applied (e.g., forensics) settings in several European nations. In a review by Butcher, Derksen, Sloore, and Sirigatti [21], it appears that countries such as Belgium, Italy, and Spain have used the MMPI or its variants for employment screening purposes. Having said that, there is very little published documentation in this area. Due to the nature of the constructs assessed by the MMPI (all versions), the resulting scores are considered to be clinical or medical in nature, and its use in personnel selection is subject to relevant employment legislation and relevance to the job [12].

4.2 COMMERCIAL INTEGRITY TESTS AND MEASURES

Table 4-1 contains a list and brief summary of commercial integrity tests that were initially included in Ones and Viswesvaran's [63] meta-analysis of integrity test validities. This table includes only those tests that were listed in *Tests in Print IX* [6], which is a comprehensive resource that describes every test in the English language that is commercially available for purchase. Efforts were made to check the currency of each test/test publisher listed in this table at the time of this writing (January 2021), but it is beyond the scope of this report to ensure the accuracy of this list due to the fluid nature of the test publisher landscape (e.g., changes in test publisher's names and/or the products they carry). Test descriptions and reviews are based on information contained in the *Mental Measurements Yearbook*, the test publisher website, and/or published research on the test. We acknowledge that there may be more recent tests of integrity on the market today. Nonetheless, we chose to focus on tests that have been subject to independent reviews published in the *Mental Measurements Yearbook*.

Table 4-1: Commercial Integrity Tests Currently in Print.

Test Name	Employee Reliability Inventory (ERI)
Test publisher (source)	Psychometrics (www.psychometrics.com)
Description	<p>This test was categorised as an overt measure of integrity in an earlier meta-analysis. A self-administered questionnaire developed by Gerald Borofsky [17], it contains 81 behaviour statements requiring a true or false response. The ERI assesses the likelihood of behaviours around seven skill areas (freedom from disrupted job performance, courtesy, emotional maturity, conscientiousness, trustworthiness, long-term commitment, safe job performance). Scoring is reported in terms of zones (one to four), which is based on the proprietary discriminant function analyses, undertaken during the validation phase, to correctly classify individuals on the basis of their ERI scores and work outcomes (e.g., accidents, unauthorised absences, disciplinary actions). The most current version of the ERI takes about 10 minutes to complete.</p>
Psychometric evidence	<p>Two reviews of the ERI were found in the Mental Measurements Yearbook [41], [72]. These reviews were based on published findings at the time. Generally, both reviewers agreed that although there is some evidence for test-retest reliability and validity, this evidence was based on small sample sizes, and should continue to be compiled using local samples on which the test is employed. The most current test manual [17] reports additional psychometric findings, some of which have also been published in the mainstream literature (e.g., Ref. [18], Borofsky, 2000). Reported test-retest estimates are based on small samples. The additional validity studies focus on examining the accuracy of each of the seven subscale scores (instead of the earlier single “R” score) in classifying individuals into criterion categories that are specific to the scale’s skill set area. Although based on small samples, 92 – 100% (validation samples) and 76 – 85% (cross-validation samples) of the cases were correctly classified.</p>
Test Name	Inwald Personality Inventory-2 (IPI-2)
Test publisher (source)	PSI Testing Excellence (www.psonline.com/assessments/the-inwald-personality-inventory-2/)
Description	<p>The initial version of this measure, the IPI, was included in Ones’ meta-analysis [61] as a personality-based measure of integrity. The IPI-2, which is a revised version of the original IPI, specifically measures characteristics and behaviours that are detrimental to high-risk professions (e.g., police officers).</p> <p>The information reported herein is based on MMY reviews of the IPI [Revised] and updated wherever possible with information about the IPI-2.</p> <p>The IPI-2 contains some new content scales that eliminate item overlap that existed in its earlier version. There are 16 content scales (Rigidity, Passivity, Admitted Illegal Behaviour, Criminal Accusations, Social Difficulties, Volatility, Anxiety, Abnormal Thoughts, Depressed Mood, Elevated Mood, Non-Conformity, Unreliability, Risk-Taking Tendencies, Irritability, Health Concerns, Substance Use) and 1 validity scale (Guardedness).</p>

Description (cont'd)	<p>The IPI contains 202 self-reported true/false questions, is accessible at a grade 5 reading level, and should take 30 – 40 minutes to complete. IPI scores are compared to a normative distribution of other public safety professionals, and those above a T-score of 60 are considered “elevated” and warrant further investigation. Candidate results are combined with Field Training Officer (FTO) predictions to provide a more complete picture of a candidate in four areas: Control of Conflict, Public Relations, Report Writing, and Overall FTO Rating. Prediction equations yield categorisations such as “likely to meet expectations” or “not likely to meet expectations” with respect to FTO performance, and an Estimated Psychologist Recommendation (“likely to recommend” and “not likely to recommend”).</p>
Psychometric evidence	<p>The IPI-2 was created with a normative sample of 89,301 pre-existing and rescored IPI records from the United States Institute for Personality and Ability Testing (IPAT) [47]. Cronbach’s alpha values for the overall IPI-2 is .52, and the individual scales range between .34 and .77. The IPAT reports that the IPI-2 had significant expected correlations (i.e., convergent validity) with the MMPI-2-RF and Personality Assessment Inventory (PAI), but the actual correlations are not listed [47]. With respect to criterion-related validity, the IPAT reports accuracy scores between 73% and 89% for the overall psychologist recommendation and FTO predictions, but does not provide other details (e.g., sample size, nature of criterion, time interval) [47].</p> <p>Some studies using the IPI (earlier version) have been published in mainstream research. For example, Detrick and Chibnall [32] reported significant associations between some IPI scales (not among the 16 revised scales) and police training performance. Studies using the IPI-2 (e.g., Ref. [56]) have used it to examine the convergent validity of other similar measures (e.g., Matrix-Predictive Uniform Law Enforcement Selection Evaluation [M-PULSE] Inventory).</p>
Test name	Personnel Reaction Blank (PRB)
Test publisher (source)	<p>PSI Testing Excellence https://www.psonline.com/assessments/the-personnel-reaction-blank-prb/</p>
Description	<p>The PRB is meant to be used by those making hiring decisions, to assess the extent to which an individual is conscientious, diligent, and has restraint. According to the publisher’s website, the test assesses an individual’s sense of well-being, compliance with rules, occupational preferences, background indicators, and has a reliability index. It contains 84 multiple-choice items and takes about 15 minutes to complete.</p>
Psychometric evidence	<p>Sundre [91] reviewed an earlier version of the PRB in the Mental Measurements Yearbook based on information available at the time. Internal consistency for Spearman-Brown was high for delinquent samples (.97 and .95 for male and females), but lower for college samples (.73 and .56 for males and females, respectively). Test-retest reliability with a sample of 26 with a 5-year delay was .56. Standard error was not reported, which is important given the intention to use the instrument in selection decisions.</p>

Psychometric evidence (cont'd)	In the manual, there was not much discussion on content validity and how the items were developed, nor on the rationale behind placing individuals on a continuum of “negative, asocial,” and “positive, conforming.” Correlations between the PRB and work performance were in the .20-.33 range, but the rating source and criterion operationalisation were not provided. Another review of the PRB [99], assuming it is based on the same data reviewed by Sundre [91], described it as being “theoretically grounded,” and with reliability and validity demonstrated in large sample studies, although they did note some gaps in the databases.
Test name	PSC-Survey ADT (<i>full name not specified in any source</i>)
Test publisher (source)	Predictive Surveys Corporation (email: marklstrand@gmail.com)
Description	The PSC-Survey evaluates alienation and trustworthiness at work. It combines three different scales into a short version. The scales are Leniency and Trustworthiness (LT), Attitudes Towards Alcohol and Drugs (AD), and Motivation, Attitude Towards Work, Supervisors, etc. (MA). The PSC is meant to replace lie detector tests asking invasive questions, which implies it evaluates honesty, but the publisher claims this is not the case. There are 100 items, scored with a correction grid. The result is two subscale scores, as well as an overall score and number of “don’t know” responses. The degree of an individual’s “virtue” is inversely related to their score.
Psychometric evidence	At the time of Jeanrie’s review [48] in the Mental Measurements Yearbook, the PSC-Survey had been released for nearly a decade. Jeanrie [48] found the lack of validity and reliability data after such a long time to be a serious deficiency and did not recommend the measure for use. Very little information is included about scale development. Only the AD scale has split-half reliability data, and the reviewer stated that Cronbach’s alpha would have been a more conservative and appropriate option (the figure is not included in the review). Validity information for the PSC-Survey is simply correlations between the three longer scales and their condensed versions: $r = .93-.96$, $n = 55-143$. Information is given for criterion-related validity, but no norms are provided for comparison, and the conclusions drawn are not convincingly supported. There is also no explanation given for why not answering a question is counted as a “wrong” answer, and the publishers claim that too many “?” responses indicate a deceptive attitude. No other published studies on the PSC-Survey were located.
Test name	PSC-Survey-Supervisory Attitude (SA)
Test publisher (source)	Predictive Surveys Corporation (email: marklstrand@gmail.com)
Description	The PSC-Survey-SA was specifically designed for selection or promotion of mid-level managers/supervisors, and measures attitudes towards supervisory positions. There are five scales that measure: attitudes towards upper management, being a supervisor, the company and business in general, towards peers, and towards subordinates. The PSC-Survey-SA consists of 60 “yes/no/?” items. To score the test, negative responses are summed, then scale totals are graphed and converted to percentile ranks on the form itself to indicate strengths and weaknesses of the applicant. The authors do not give an explanation or description of how to interpret scores other than that lower scores are better. The PSC-Survey-SA takes around 15 minutes to complete.

<p>Psychometric evidence</p>	<p>Pearson [64] and Witta [105] reviewed the PSC-Survey-SA in the Mental Measurements Yearbook. Overall, the authors agreed that there was little evidence to support using the PSC-Survey-SA to make selection decisions.</p> <p>Split-half reliability for the PSC-Survey-SA was .74, and no other reliability data were provided. There is some evidence for concurrent validity, where a sample of 15 individual’s PSC-Survey-SA scores were compared to supervisor ratings at .72. A study on predictive validity recorded PSC-Survey-SA scores for 42 men and 10 women who were recently hired. Supervisor ratings were collected six months later, but a correlation coefficient is not reported. Witta [105] hypothesises that the results are skewed because only those with low PSC-Survey-SA scores were likely hired. There is a standardised sample of 72 females and 97 males, but there is no explanation as to how the percentiles were derived. Pearson [64] (1998) suggested a number of other well-validated leadership surveys to use instead (e.g., Leader Behaviour Analysis II from Blanchard Training & Development, Inc.). Witta [105] was more optimistic and stated that the PSC-Survey-SA could be useful in screening applicants for first-line supervisory positions.</p>
<p>Test name</p>	<p>Reid Report (RR)</p>
<p>Test publisher (source)</p>	<p>PSI Testing Excellence (https://www.psionline.com)</p>
<p>Description</p>	<p>According to an early review by Brodsky [19], the RR underwent 19 revisions in the first 30 years of its creation, and has been given to over 5 million people, mostly in low-level positions (e.g., labourers, clerks). The test is intended for screening in positions where honesty and integrity are critical to the work. The test booklet contains four parts, with a total of 320 items. The first part, the Integrity Attitude Inventory (IAI), consists of 80 yes/no items which are scored for use. The other parts assess work history, drug and alcohol abuse, and antisocial history; these are optional. Only 70 of the 80 questions in the IAI are scored. There are meant to be four types of final evaluations: recommended, qualified, not recommended, and no opinion. However, the manual only refers to “recommended or not”, with a cut-off score of 49-50 and it does not explain how the cut-off was determined. Approximately 75% of applicants receive a ‘recommended’ evaluation. The test should take 15-30 minutes depending on the version [25].</p>
<p>Psychometric evidence</p>	<p>The RR has been reviewed multiple times in the Mental Measurements Yearbook, and a synthesis of these reviews is provided here. The reviewers seem uncertain as to the usefulness of the RR due to missing information. At the time of the most recent review, the only psychometric data available was for the IAI [34]. The manual states that 18 internal consistency studies have been completed, but only 14 are listed, with reliabilities ranging from .71 – .86. In addition, Cronbach’s alpha was provided in some cases, which is inappropriate for the dichotomous nature of the items.</p> <p>There is no discussion of test development, or evidence to support the four-factor structure claimed by the developers, which does not include honesty or integrity. A predictive validity study [27] found RR scores correlated with honest behaviour, but the RR scores correlated even more strongly with the tendency to impress others and to deny undesirable qualities.</p>

Psychometric evidence (cont'd)	More recent evidence [95] suggested the RR was not valid in predicting any type of job performance. Murphy [57] states that an APA task force strongly criticised the use of categorical/dichotomous scoring in integrity testing, which is the foundation of the RR.
Test name	The Stanton Survey (SS) and the Stanton Survey Phase II
Test publisher (source)	The Plotkin Group (https://www.plotkingroup.com)
Description	<p>The SS provides information on attitudes, values, beliefs, and information on past dishonest behaviour. It is based on seven concepts: General Theft, Opportunism, Employee Theft, Leniency, Employee Discounting, Pervasiveness, and Association. There are 74 yes/no and multiple-choice questions, with options to write further explanation. The SS also includes requests for personal information on schooling, addresses for the past five years, and past employment. The SS is still scored by the publisher at this point, so no information on scoring or norms was provided, but statistical data is available from the publisher upon request. It is known that attitudinal items are summed for a numerical base score (NB), and admissions of wrongful acts (AS) are categorised as petty, marginal, or severe. NB and AS scores are then summed for a Stanton Score which places applicants into Low, Marginal, or High-Risk categories, indicating levels of dishonest behaviour.</p> <p>The SS Phase II contains 36 “attitude” items that are agree/disagree, with some factual items with 2-10 options such as “How much money have you taken (stolen) in your life?” It is scored the same way as the SS.</p>
Psychometric evidence	<p>Overall, the three reviewers in the Mental Measurements Yearbook felt that evidence for the SS and Phase II did not satisfactorily support the prediction of stealing/cheating behaviour of employees [38], [43], [104]. The only reliability data provided is a Kuder-Richardson’s coefficient of .91 (internal consistency). A few validity studies were described, but the operationalisation of “criminal” groups was based on admissions of previous misconduct, or polygraph testing. The generalisability of the SS across 10 occupational groups was tested by using the NB as a predictor and AS the criterion with a sample of 5,325. The validity coefficient was .70, but there is no information as to how this was calculated. Criticisms included the high potential for faking, and the invasiveness of the biographical data and some questions. A later review [100] described the Stanton Survey as well-researched with demonstrated criterion-related validity in security samples. The report also indicated that the measure is reliable and does not have adverse impact.</p>

The tests reviewed in Table 4-2 are largely North American in origin. Therefore, we have included a summary evaluation of several integrity tests that are known and used in some Scandinavian organisations for personnel selection (see Table 4-2). Some of these tests are DNV certified. DNV-GL, a Norway-based company involved in diverse quality assessments and risk management, evaluates commercially available tests in line with the guidelines for test reviews developed by the European Federation of Psychologists’ Association [35]. A DNV certificate indicates that the test has been reviewed and deemed to be a valid assessment.

Table 4-2: Integrity Tests Used in Scandinavian Organisations.

Test Name	MAP-X
Test publisher (source)	Assessio (www.assessio.com/en/mapx-2)
Description	<p>MAP-X is a self-assessment test based on the test publisher’s Five Factor Model (FFM) test called MAP. The MAP test and MAP-X have the same 200 items, 40 per scale, and 5 sub-scales per FFM dimension. MAP-X is an expanded evaluation of MAP test scores, using the alternative DSM-V Model for Personality Disorders [3], [4], [9], [83].</p> <p>The MAP-X test is administered online through the publisher’s platform called Ascend. The test can be delivered via computer, iPad, or smartphone. Test results are available both as test scores for comparison (ranking) of several candidates, and as narrative reports for individual analysis of test results. The test takes about 30 minutes to complete.</p> <p>According to the MAP-X technical manual, the test is supposed to not only map the individual’s test scores to the five dimensions of the FFM (including facets), but also identify an individual’s risk tendencies (also called derailers) correlated to their personality. These tendencies may lead to extreme or counterproductive work behaviours. The results may also help to identify how positive personality aspects may become toxic or critical derailers under extreme or stressful conditions. The test results are further meant to be useful in recruitment and selection, as well as in the development of coworkers, and to assess who might be suitable for certain tasks and positions [9].</p> <p>MAP-X is also described as a measure of the extreme ends of the FFM variables. Therefore, this MAP-X test actually has ten new scales, one in each of the extreme low ends of the FFM factors and one in each of the extreme high ends of these factors (e.g., for Extraversion, extreme low end: Withdrawn; extreme high end: Unrestrained). The scales are presented in Table 4-2, along with the normal FFM factor in between. The MAP-X scales should be understood as measures of the dysfunctional and maladaptive dimensions, on the extreme ends of the FFM. For further descriptions of each scale please see the technical manual [9].</p>
Psychometric evidence	<p>Since both tests have the same items, the internal reliability for MAP and MAP-X is optimal ($> .80$), with a robust factor structure. The manual reports a rather good convergent and discriminant validity with other tests. The MAP-X scales were positively correlated with the test scores of MINT, published by the same test publisher Assessio, as well as with the Honesty-Humility scale (from the HEXACO inventory).</p> <p>MAP-X also correlates well with the PID-5, a measure of the DSM-V alternative model of personality disorders ([3], [4]), demonstrating that MAP-X measures the dysfunctional and maladaptive dimensions on the extreme ends of the FFM. For example, MAP-X Emotional Stability correlates negatively with PID-5 Negative Affectivity ($-.82$), MAP-X Extraversion correlates negatively with PID-5 Detachment ($-.70$) and MAP-X Conscientiousness correlates negatively with PID-5 Disinhibition ($-.69$).</p>

<p>Psychometric evidence (cont'd)</p>	<p>As reported by the test publisher, MAP-X scales of Agreeableness, Conscientiousness, and Emotional Stability held moderate to strong negative correlations (-.14 to -.58) with each of the three dimensions of the Dark Triad, that is scales of Machiavellianism, Psychopathy and Narcissism as measured by the self-report test <i>Dark Triad Dirty Dozen</i> [9].</p> <p>CWB was moderately and negatively correlated with the Agreeableness scale, as well as the Conscientiousness and Emotional Stability scales. Work engagement and self-rated Job Performance was positively correlated with all the big five scales [9], [83].</p> <p>The manual reports more correlations with other tests. For example, the correlations between MAP-X and the Hogan Development Survey (HDS) [9] are a bit more complex, which is partly explained by the fact that the tests are based on different taxonomies. The HDS is based on DSM-IV categorical approach, while the MAP-X is based on the DSM-5 alternative personality model ([3], [4])</p> <p>More empirical research is needed, though, to report more about MAP-X's actual predictive validity in predicting job performance in actual selection procedures and/or development assessments. Since MAP-X is a rather new measure, additional research findings are likely to be published in the future. At the time of this writing, the MAP-X version of the test did not have a DNV certificate.</p>
<p>Test name</p>	<p>MINT (Measuring INTEgrity)</p>
<p>Test publisher (source)</p>	<p>Assessio (www.d-teck.com/products/mint-assessio)</p>
<p>Description</p>	<p>According to the test manual, MINT is a personality-based integrity test that purports to predict job performance and CWBs [79]. The theoretical model for this test is based on the Big Five personality dimensions of Conscientiousness, Agreeableness, and Emotional stability. The two frequently used dimensions of Interpersonal orientation and Task orientation in categorising different forms of CWBs are also used for a comprehensive model for this test.</p> <p>Each personality dimension is measured with 20 items, and the test can be administered online in an unproctored setting. There is no time limit. A weighted combination of the three personality traits makes up the main integrity or the overall score. Scores are presented on a 10-point scale and are also provided for the two factors of Interpersonal Orientation and Task Orientation. The manual specifies that scores on the two factors are “independent,” thus it is possible to obtain any combination of these factor-scores without the overall integrity score being affected. The manual recommends that the overall integrity score should serve as the basis for selection decisions, because the scores on the factors do not provide information about risks for CWBs. The test manual reports thorough analyses of item development, reliability, and convergent validity for the test, some of which is included below and focusses on employee samples and the overall integrity score [79].</p>

<p>Psychometric evidence</p>	<p>The test manual [79] reports a correlation of $r = .25$ (operational validity estimate of .36) between the Integrity score and job performance among traffic controllers in Sweden ($n = 47$), as evaluated by their managers. In the same sample, a self-reported job performance rating correlated quite high with the integrity score ($r = .48$, operational validity estimate of .67), but not for self-reported job satisfaction ($r = .05$). Correlations between the MINT Integrity score and job performance are also reported for a Norwegian sample of workers in a grocery store chain ($n = 101$, $r = .19$) and a Norwegian sample of warehouse workers and truck drivers ($n = 68$, $r = .13$). The operational validity estimates for these two samples in Norway were .26 and .18, respectively. For these reported correlations, it is unclear if a concurrent validity strategy was used, or if the reported correlations were based on a predictive validity strategy with a time span between test result obtainment and criterion measures. Some of these studies were used for test-retest analyses, in which a time span of 30 days is reported between the testing. This could mean that this was the maximum timespan between test results and criterion measures for some of the studies.</p> <p>Studies using criterion measures of CWBs are of especial interest for integrity testing. There is one empirical finding on CWBs in the test manual, based on the Norwegian sample of grocery store workers ($n = 101$). Scrutinising the details, CWBs were measured by a supervisory rating scheme consisting of nine questions. These questions were directed at both job performance and CWBs, but it is not reported how many items were devoted to each criterion. Only two examples of CWB-items are reported: “showing up late to work” and “non-job-related use of the internet.” Thus, the exact content of the CWB evaluation completed by supervisors is unknown. The final sample used for the analysis of MINT and CWB were $n = 44$, partly because of missing data. The reported correlation between the integrity score and supervisor-evaluated CWBs was $r = -.28$ ($n = 44$), where the reported operational validity estimate were -.35.</p> <p>The MINT is a DNV-GL certified test for use in personnel selection, and therefore holds a certificate that claims that Norwegian and Swedish versions of the test are valid assessments (i.e., the test predicts what it claims to predict [i.e., job performance and CWBs], [33]). But unless the DNV-GL has secured more information on MINT’s predictive validity towards CWB than what is reported in the test manual, MINT’s certificate should be considered problematic.</p>
<p>Test name</p>	<p>Squares</p>
<p>Test publisher (source)</p>	<p>Aon Assessment Solutions, formerly cut-e (www.aon.assessment-solutions.com)</p>
<p>Description</p>	<p>Squares is described as an integrity test that aims to identify applicants with a likelihood of CWBs [7]. According to the test manual, the theoretical background of Squares was based on an identification of six situational aspects (distraction, ambiguity, boredom, indifference, opportunism, superficiality) believed to increase probability of CWBs. These aspects are each mapped onto respective psychological attributes (discipline, conscientious, cautious, empathetic, honest, and reflective) hypothesised to determine individuals’ susceptibility to situational forces. These attributes are further categorised into two overarching factors, impulse control and ethical awareness, which are deemed to reflect a trait called trustworthy.</p>

<p>Description (cont'd)</p>	<p>Squares is administered online as an unproctored test. Six items (e.g., I often and willingly help others) measure each facet, and test takers are asked to rate each behaviour by comparing themselves to other people. The rating scale is a sliding bar that test takers can move towards the scale's three anchors (<i>holds less true for me than for others, holds equally true for me and for others, holds more true for me than for others</i>). Unweighted summed scores from each facet are used to yield scores on the two factors and the overall trustworthy scale. Based on a combined German-speaking and English-speaking sample ($n = 335$), the average completion time for the test was 7.5 minutes. In Norway, Squares has been used in suitability evaluations of police students for some time.</p>
<p>Psychometric evidence</p>	<p>In the test publisher's report [7] (cut-e, 2013), Cronbach's alphas for the facets and factors were .71-.90, whereas the total trustworthy scale had an alpha of .92. Intercorrelations of the facets were between .17 and .52. Findings from a concurrent validation study with 33 participants in Germany, which correlated scores on Squares with one-hour long structured interviews scoring each facet using a 10-point scale (0 = <i>very low</i>, 5 = <i>average</i>, 10 = <i>highly pronounced</i>), found correlations in the range of .60-.70.</p> <p>Further psychometric evidence comes from an independent study conducted by one of the authors of this chapter [84], [85]. This was based on data gathered from police applicants during the Norwegian police selection process in Spring of 2014, and from police students attending the Norwegian Police University College in the years 2014-2017. To become an authorised police officer in Norway, a three-year educational program resulting in a Bachelor's degree is required.</p> <p>In the selection phase, Squares was among the psychological tests administered to applicants/students. Three criterion measures were examined:</p> <ol style="list-style-type: none"> 1) Outcomes of the security clearance process (passed or failed the personnel security evaluation); 2) Theoretical and practical performance ratings in the educational period; and 3) Completion of the police-degree within the expected time range for the police education. <p>Security clearance outcomes (coded 0 = fail and 1 = pass) for the applicant sample ($N = 1261$), 47 of whom did not pass the security evaluation, did not show any significant correlation with Z-scores of the Squares' factors of Impulse control ($r = -.03$) and Ethical awareness ($r = .02$), or the total Trustworthy score ($r = -.02$). In a sample of 619-664 students, average grades in five main study areas (uniformed police service, methodology, policing and society, crime prevention, and criminal investigation) were also correlated with the total Squares' scores (i.e., trustworthy). However, Squares scores were found to be unrelated to grades in all study areas ($r = .04$, uniformed police service; $r = -.04$, methodology; $r = .02$, policing and society; $r = .03$, crime prevention; $r = .01$, criminal investigation).</p>

Psychometric evidence (cont'd)	<p>Performance in a 10-month on-the-job training period, completed in the second year of study, was evaluated using a five-point Behaviourally Anchored Rating Scale (BARS) that was developed to measure the same six competencies (interaction, openness and inclusiveness, maturity, power to act, integrity, and analytical capacity) that the students had been evaluated on in the selection process. Each competency contained five items that instructors provided ratings on. Of the 160 instructors that were asked to participate, 107 returned the rating scale (response rate of 67%).</p> <p>The resulting correlations between Squares trustworthy scores and competency ratings were found to be low and non-significant ($r = -.15$, interaction; $r = -.14$, openness and inclusiveness; $r = .08$, maturity; $r = -.08$, power to act; $r = .01$, integrity; $r = -.05$, analytical capacity).</p> <p>Squares (factors and total) Z-scores were also correlated with a dichotomous variable indicating if the student had completed the educational degree in the expected time (coded 1) or not (coded 0). Of the 668 students in the sample, 94 students did not complete the program within the expected time due to failure on one or more exams taken in the last spring before graduation. There were significant (point-biserial) correlations, although in the wrong direction, meaning that higher Squares scores increased the probability of not completing the degree in the expected time ($r = -.14$, impulse control; $r = -.10$; ethical awareness, $r = -.13$, trustworthy).</p>
Test name	Understanding Personal Potential (UPP) Personality and UPP Integrity
Test publisher (source)	Psykologisk Metod AB (Sweden) (www.psykologisk-metod.se)
Description	<p>This test is a measure within the commercial test battery called UPP, consisting of the UPP Personality (personality test), the UPP Problem Solving (general intelligence), the UPP Screen (short personality test), and the UPP Integrity (claimed by the test publisher to measure integrity).</p> <p>The UPP Personality has 15 scales, measuring up to 29 dimensions including the factors of the Five Factor Model (FFM; Extraversion, Emotional Stability, Openness, Agreeableness, Conscientiousness). Other scales are aimed at measuring other personality aspects and important dimensions within occupational life, such as Willingness to Work, Control orientation, Initiative, Self-Appreciation, Emotional Intelligence, etc.</p> <p>The UPP Integrity is a set of eight dimensions taken from the full personality test called UPP Personality, where some of the scales are reverse-scored and labelled with new names. UPP Screen also consists of a set of dimensions taken from the full personality test UPP Personality, but these are not the same set of scales as UPP Integrity.</p>

<p>Description (cont'd)</p>	<p>The dimensions in the UPP Integrity test are:</p> <ul style="list-style-type: none"> • Agreeableness • Conscientiousness • Emotional Stability • Emotional Intelligence • Passive Aggression (inversed Willingness to cooperate) Self-Appreciation (Narcissism) • Psychopathy (inversed Benevolence) • Machiavellianism (inversed Unselfishness) <p>The UPP Personality (and UPP Integrity) tests also include questions and algorithms indicating data quality of the responses of the test taker. This consists of analysing the test taker’s mood and attitude, as well as response behaviour, in the test situation. The UPP Personality test includes a statistical method for correction of faking (or impression management). This is based on using measures of two different social desirability scales. The overt measure is of the Crowne-Marlowe type, while the covert measure is based on items which are not clearly socially desirable indicators but are selected since they correlate strongly with the overt scale [80]. The correction method used is controversial, though, due to generally weak support of statistical methods for correction of faking within the science community.</p> <p>Test scores of each dimension are presented with and without the correction for faking, giving the test administrator indication of the value of the difference (or the size of faking). Their own studies have shown that faking differs across different variables. Therefore, the faking correction method in the test is adapted accordingly. The test also has scales used for simultaneous validation, which in the manual is called “proxy-validation.” The test includes questions and scales related to work which leads to scores on several work-related dimensions such as Willingness to Work, Job interest, Job Satisfaction, Balance work/life, etc. Scores on these work-related dimensions are used as criteria for a simultaneous validation of all the other personality-related dimensions of the test. Therefore, by using a set of general proxy criteria (which is the test taker’s responses to certain work-related dimensions within the test) instead of using specific real-life work criteria (which is usually hard to collect for every test taker), this type of simultaneous validation (or “proxy validation”) results in a score which indicates how valid the test results are on predicting performance in working life (or general job performance).</p>
-----------------------------	--

<p>Description (cont'd)</p>	<p>The test is administered online. Test results are presented as a test report of choice per individual (only psychometric scores, brief narrative, long narrative). There are also special feedback versions of the report aimed for the test taker, if necessary. Data files for groups of test results are possible to retrieve. All test scores are presented according to available standardised norm groups. Test scores of each dimension are also presented with and without the correction for faking, giving the test administrator indication of the value of difference. According to the test publisher the test is available in several language versions (Swedish, English, Norwegian, Russian, Spanish, Arabic, Finnish, Danish, Tigrinja, Dari).</p> <p>The test publisher claims their online tests have undergone accessibility analysis and adaptations to minimise eventual discrimination, according to a review they initiated which was conducted by a Swedish national accessibility organisation [37].</p>
<p>Psychometric evidence</p>	<p>This section focusses on psychometric evidence of the UPP Integrity Test, as is reported by internal studies referred to in the manual [66], [81].</p> <p>The Unselfishness scale is the inverse of Machiavellianism in the UPP Integrity test. The Technical manual of UPP ([65], [66], [67]) reports that the Machiavellianism scale is well-validated, because it has high internal reliability (.90) and correlates highly (.90) with a modernised version of a Narcissism scale [29].</p> <p>In a recent study by Westlin [102], [103], the Psychopathy and Machiavellianism scales predicted CWB significantly ($r = .37$ and $r = .26$, respectively), while Narcissism failed to predict CWB in this study. However, Narcissism scale was found to consistently have predicted CWB in other internal studies, and the UPP Narcissism scale showed convergent validity with Raskin and Terry's (1988) Narcissistic Personality Inventory –16 (uncorrected $r = .63$, corrected $r = .80$; Ames et al., 2006), the Narcissism scale is still included in the UPP Integrity test.</p> <p>The passive-aggression scales in UPP Integrity and the “Leisurely” scale of the Hogan Development Survey (HDS; [45]) are explained to be subclinical versions of the same construct and are moderately correlated with each other(-.42 uncorrected and -.66 corrected) [66], [81], [102].</p> <p>An Emotional Intelligence measurement is included in UPP Integrity based on rather questionable support from the scientific literature, but it is still considered a factor of great importance to working life in associated reports and results [65], [66], [67], [82]. Correlations of the UPP Emotional Intelligence measure with other such scales was found to be in the ranges of .36 – .54 (uncorrected) and 0.5 – .68 (corrected). In a criterion validity study, Westlin (2019) reported that emotional Intelligence and OCB were correlated ($r = .36^{**}$). The UPP (and assumingly all the tests contained within it) is DNV certified [33].</p>

4.2.1 Summary

The objective of this section was to provide insight into commercially available measures of integrity. Of the 25 tests included in Ones et al.'s [63] meta-analysis, only eight (see Table 4-1) are currently available (as of January 2021). In addition, we included a review of several integrity tests (see Table 4-2) that tend to be used in Scandinavian organisations. Of note, a recent assessment called the Wartegg Drawing Completion Test – Crisi Wartegg System, which is a performance-based personality measure utilising an assessment technique similar to projective tests such as the Roscharch [26], was also considered but not included here because it does not claim to assess integrity.

As indicated through our review, the composition of each measure differs, with some focussing on personality traits, and others on attitudes, values, past experiences, or a combination of these varied constructs. The psychometric properties of each measure also vary, with some more psychometrically sound than others. Should an organisation wish to use a commercial measure of integrity in personnel selection, it should examine the generalisability of any existing psychometric evidence that is deemed sufficient by comparing its local operational context and requirements against those in which the evidence was first gathered [87]. Alternatively, efforts to gather local validity evidence (i.e., within one's own organisational context) may be undertaken to support the use of a measure in personnel selection [87].

4.3 EMERGING APPROACHES TO ASSESSING INTEGRITY

It is clear from the other chapters in this report that technology has influenced personnel selection and assessment in many ways. This section examines the extent to which technological advances have facilitated innovative approaches to assessing integrity. In describing emerging technology trends in staffing/assessment, Jones and Dages [50] used the terms platform-centric and information-centric to categorise some of these trends. They described platform-centric approaches as the use of computer hardware or software, and/or informational technology infrastructure required to support an assessment. Consequently, we view platform-centric approaches as entailing the administration of traditional assessment tools (e.g., paper-based inventories, in-person interviews, role play exercises) using technology (e.g., computer or online inventories, web-enabled interviews). Information-centric approaches, according to Jones and Dages [50], refer to the use and processing of technology-based information. As a result, we see information-centric approaches as using and processing information that is generated on web-enabled platforms such as Social Networking Sites (referred to as social media here) to assess some construct. Examples included Facebook, MySpace, Twitter, YouTube, and others. Our review of the literature around integrity assessments is organised into these two categories.

4.3.1 Platform-Centric Approaches

Dages, Zimmer, and Jones [28] examined the psychometric properties of a shortened version of the Reid Report, an overt measure of integrity, using mobile technology (e.g., smartphones, tablets). The Reid Report (see Table 4-1) is considered to be a traditional assessment of integrity because it was included in Ones, Viswesvaran, and Schmidt's [63] meta-analysis, along with other traditional integrity measures. As part of an organisation's pre-employment screening process, Dages et al. [28] created a shortened and unproctored version of the Reid Report, assessing integrity attitudes and admissions of counterproductive work-relevant behaviours. Applicants completed this measure by accessing a web-link to this version of the assessment using a desktop/computer, tablet, or smartphone. The authors found estimates of reliability and validity were comparable across the type of devices used.

Jones, Brasher, and Huff [49] conducted a series of studies to examine the Applicant Potential Inventory, which was adapted from a traditional integrity assessment, the Personnel Selection Inventory (see Table 4-1). The Applicant Potential Inventory represented a modified version of the Personnel Selection Inventory (i.e., shorter, lower reading level, and fewer response options), specifically developed for administration using an interactive voice recognition (IVR) system (i.e., automated telephone delivery of questions, recording of touch-tone key responses, and scoring of responses). Findings from six studies provided evidence for strong convergent validity between the IVR-based Applicant Potential Inventory and the paper-pencil Personnel Selection Inventory, comparable reliability and validity coefficients between the IVR-based Applicant Potential Inventory and paper-pencil version of this inventory, no adverse impact or group differences using the IVR-based inventory, as well as concurrent and predictive validity of the IVR-based inventory in predicting supervisor ratings of performance in two different employee samples.

Aguinas, Henle, and Beaty [1] discussed the potential use of virtual reality technology (another platform) in the assessment of certain constructs, including integrity. They explained that because virtual reality allows users to more fully immerse themselves in a particular scenario or task, it can increase the fidelity of a selection tool by creating a more realistic representation of the job-related requirement. This can increase the validity of the selection tool. However, the high costs associated with such technology prohibits its use, and likely explains why we were unable to find published research on the use of such technology in integrity assessment, or in personnel selection at writ large.

4.3.2 Information-Centric Approaches

The use of information-centric approaches in organisational screening is on the rise. Clark and Roberts [23] provided a summary of various studies on the use of this practice, revealing that 50% of HR managers reported running an internet search on applicants [106]. The primary purpose of such practices is to uncover negative or CWBs (e.g., illegal drug use, slandering employers or coworkers) in candidates [31], [86]. Berkelaar [13], [14] used the term “cybervetting” to refer to the practice of employers gathering information about prospective employees using informal, online sources to inform selection decisions. Hence, this practice involves using types and sources of information that deviate from traditional measures. As Clark and Roberts [23] described, social media provides organisations with inexpensive and readily available information that is used to conduct quick online background checks on applicants. Traditional background checks are typically expensive, dictated by the requirements of the job, and reserved for serious candidates.

With the rise of information-centric approaches in selection practice, research into the validity of such approaches is slowly accumulating. While some research concerns the assessment of various constructs (e.g., personality, intelligence), it provides insight that is relevant to the assessment of integrity. For example, Newness, Steinert and Viswesvaran [59] shed light on the extent to which the type and amount of information posted on social media is indicative of an individual’s personality. They were particularly interested in predicting the posting of inappropriate information (e.g., semi-nude or nude photos, comments about illegal drug use, student code violations, or alcohol consumption) in a college student sample. They found that integrity-relevant personality factors, Conscientiousness and Honesty-Humility, were the most strongly and negatively correlated with inappropriate postings. Becton, Walker, Schwager, and Gilstrap [11] examined the validity of inappropriate social media content, specifically its relationship with CWB, workplace accidents, and alcohol consumption. They found that inappropriate social media content (coded in terms of profane language/sexual content and alcohol/drug consumption) was unrelated to CWB or accidents; however, they found small positive associations between inappropriate social media content and self-reported alcohol consumption and heavy drinking.

Van Iddekinge, Lanivich, Roth, and Junco [96] examined the predictive validity of recruiters' ratings of ten job-related attributes (e.g., interpersonal skills, adaptability, creativity, integrity, intelligence) as well as overall suitability of prospective applicants' social media profiles. They found ratings were unrelated to supervisor ratings of performance, turnover, or user-reported turnover intentions. They also compared recruiters' suitability ratings based on users' sex and ethnic group and found that female and white user profiles were given higher suitability ratings. This is primarily due to ratings on conscientiousness-related attributes (i.e., integrity, work ethic, dependability, and maturity).

Kluemper and Rosen [52] examined whether social media information can be reliably and accurately evaluated or rated to identify personality traits and job-relevant information. Student raters were asked to review six social media user profiles, and to rate each user's personality using an adjective checklist measure of the Big Five, and to estimate the users' intelligence and Grade Point Average (GPA). They found high agreement between raters in their ratings of all constructs, as well as high internal consistency and reliability estimates for all personality factors. Kluemper, Rosen, and Mossholder [53] later replicated this finding over two studies using a larger sample of user profiles and found acceptable inter-rater agreement between three raters in their ratings of personality and intelligence based on social media user profile information. In addition, they examined the convergent and criterion-related validity of these ratings. With respect to convergent validity, they concluded that personality ratings based on users' social media user profile information had small-to-moderate associations with self-reported personality assessments, which were reportedly in the same range as correlations between self- and other-report ratings of personality (see meta-analysis in Ref. [24]). With respect to criterion-related validity, they discovered that personality ratings based on social user profile information predicted supervisor ratings of users' job performance (employed sample) and academic performance (student sample). These ratings also had incremental validity over user- (i.e., self-) ratings of personality, with some Big Five factors accounting for more incremental variance than others.

4.3.3 Implications

A number of the findings with respect to emerging approaches to assessing integrity are promising and may have theoretical relevance. However, much of the literature cautions against the use of information-centric approaches in selection, particularly social media. They emphasise the need for additional evidence using larger and varied samples across organisational contexts and factoring in other considerations (discussed below).

4.3.3.1 Theoretical Relevance

With respect to the theoretical relevance of social media information, Kluemper and Rosen [52] note that social media generally provides information on the typical behaviours of an applicant. This is in contrast to the maximal performance indicators captured through application forms, resumes, and other selection assessments. Consequently, social media information is likely to reflect a more accurate representation of the typical behaviours of an applicant. This is in turn more likely to predict typical performance in a job, or at least add to the validity of assessment tools focussed on maximal performance. Becton et al. [11] cited interactionist theories [20] and the theory of reasoned action [2] to support the usefulness of social media as a proxy for the user's social environment, and reflective of the user's attitudes, interests and values. Citing Funder's [36] realistic accuracy model, Kluemper et al. [53] explained that the richness of social media allows raters to form a schema or representation of the user. This has the potential to enhance rating accuracy, particularly in the evaluation of personality traits. Therefore, there appears to be plausible theoretical justification for considering the use of social media in assessing integrity. However, there are various practical and legal factors that limit the use of such information-centric approaches.

4.3.3.2 Applicant Reactions

A small body of research has investigated applicant perceptions on the use of information-centric approaches. For example, one qualitative study (e.g., Ref. [14]) found that there is a shift in the implicit expectations of employers and applicants in the use of digital information. The author found that even though both groups regarded cybervetting as inevitable, applicants had concerns about transparency in the use of online information. They were particularly worried about misinterpretation and use of possibly inaccurate information. Employers, on the other hand, expect prospective applicants to take responsibility for the management of their online presence. Using an experimental manipulation of a job screening process, Stoughton [88] found that groups led to believe that their social media information would be used to screen their applications felt an increased invasion of privacy and perceived the process to be less procedurally fair. This lowered their perceptions of the organisation's attractiveness. These findings were replicated in another study by Gomez [39]. In this instance, social media screening was seen as an invasion of privacy and the selection process was viewed as procedurally unfair, regardless of whether it occurred before or after a job offer had been made. Another study by Madera [55] concluded that the use of social media, whether in an organisation's selection or promotion process, impacted applicants' perception of procedural fairness and their job pursuit intentions.

4.3.3.3 Job-Relevance

The risks associated with establishing the job-relevance of social media information has been repeatedly mentioned throughout the literature surveyed. This concerns the establishment of the validity and reliability of social media information in predicting job-related outcomes [30]. In addition, Davison, Maraist, Hamilton, and Bing [31] also discuss the irrelevancy and inaccuracy of social media. Organisations perusing social media for job-related information also have access to information about the user's religious beliefs or sexual orientation. This is information that is not typically solicited in traditional selection processes because it violates human rights legislation in some nations. Information on social media may also be intentionally falsified by the user (e.g., embellished education or job-related experience) or contain false negative posts by a third-party to intentionally tarnish the user's reputation. For all these reasons, it is more difficult for organisations to legally defend selection decisions that incorporate the use of information from social media.

4.3.3.4 Lack of Standardisation

Brown and Vaughn [16] discussed the lack of standardisation in the perusal and/or collection of social media information on the part of the organisation. In addition, they look at variability in the type and amount of information that is contained in a user profile. As such, the use of social media information resembles an unstructured measurement approach, which is considered to be less psychometrically sound compared to more structured approaches and increases the potential for legal challenge. Although it is possible to standardise the collection of social media information, the effort may be ineffective, as demonstrated in Schneider's [77] research. This study attempted to standardise collection of social media information by identifying specific indicators of the Big Five personality factors and integrity, and developing a rating protocol to evaluate the information. However, the resulting ratings demonstrated low inter-rater reliability, and were either uncorrelated or had low correlations with traditional measures of these same attributes. This research also demonstrated that close to 60% of employed social media users would likely refuse a request to access their user profile by an employer. As described in Davison et al. [30], there may be generational and racial differences in users' acceptance and use of social media information for hiring purposes. Together, these findings suggest that even if it was possible to develop a psychometrically sound structured approach to assessing social media information, the lack of accessibility across user profiles would limit the effectiveness and legal defensibility of this approach.

4.3.3.5 Legal Issues

It is evident that there are many legal concerns about the use of information-centric approaches (e.g., social media information) in selection. Key among these are privacy and discrimination (see Ref. [30] for fuller coverage). There is also variation across nations in regulations concerning the use of data/information. For example, the European Union has stricter regulations compared to the United States. But individuals in many countries around the world expect a right to privacy. Consequently, organisations choosing to use social media information in their hiring decisions may find themselves in violation of their nation's privacy laws. Davison et al. [30] and Brown and Vaughn [16] also discussed employment equity and disability acts and/or laws. These laws/acts ensure that selection decisions are not made on the basis of an applicants' demographic characteristics (e.g., race, sex, age), preferences (e.g., political affiliation, religious beliefs), and/or health status (e.g., disability, mental health). Yet, such characteristics may be easily available or deciphered through information on social media, placing an increased onus on the organisation using such information to demonstrate that it has not unfairly discriminated against an applicant or unfairly dismissed an employee.

4.4 CONCLUDING REMARKS

The objective of this chapter was to provide insight into existing assessments and emerging approaches used to assess integrity. The chapter provided a general overview of overt and covert approaches, and a brief description of other specific measures (e.g., polygraph tests, MMPI) that have been used in pre-employment screening in some sectors to identify honest candidates or those at risk for counterproductivity. It also included a list and brief review of currently available commercial measures that claim to assess integrity. Finally, it reviewed the literature with respect to emerging approaches used to assess integrity, including a discussion about the implications of using such measures in selection. It is hoped that this comprehensive coverage of approaches and measures used to assess integrity will be useful to those seeking to identify options or measures to assess integrity in personnel selection.

4.5 REFERENCES

- [1] Aguinas, H., Henle, C.A., and Beaty Jr, J.C. (2001). Virtual reality technology: A new tool for personnel selection. *International Journal of Selection and Assessment*, 9(1-2), 70-83.
- [2] Ajzen, I., and Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- [3] American Psychiatric Association (2013a). *Diagnostic and statistical manual of mental disorders (5th Ed.; DSM-5[®], pp 761-781)*. Washington, D.C., U.S.A.: American Psychiatric Association. ISBN 978-0-89042-555-8.
- [4] American Psychiatric Association (2013b). *Alternative DSM-5 model for personality disorders*. FOCUS Spring 2013, vol XI, No.2. Washington, D.C., U.S.A.: American Psychiatric Association. Retrieved 15 Mar, 2021, from <https://focus.psychiatryonline.org/doi/10.1176/appi.focus.11.2.189>
- [5] Ames, D.R., Rose, P., Anderson, C.P. (2006). The NPI-16 as a short measure of narcissism. *Journal of Research in Personality*, 40(4), 440-450.
- [6] Anderson, N., Schlueter, J.E., Carlson, J.F., and Geisinger, K.F. (2016). *Tests in print IX: An index to tests, test reviews, and the literature on specific tests*. Buros Center for Testing.

- [7] Aon Assessment (2020). Personality questionnaire to measure integrity (squares). Retrieved from www.aon-assessment-solutions.com/nc/us/details/personality-questionnaire-to-measure-integrity-squares/
- [8] APA Task Force (1991). Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions: An A.P.A. Task Force report. Washington, DC: American Psychological Association.
- [9] Assessio (2020). MAP-X Technical Guide. Stockholm, Sweden: Assessio International AB. Retrieved 12 Mar, 2021, from https://support.assessio.com/hc/en-us/article_attachments/360008493437/MAP-X_techmanual.pdf
- [10] Baert, S. (2018). Facebook profile picture appearance affects recruiters' first hiring decisions. *New Media & Society*, 20(3), 1220-1239.
- [11] Becton, J.B., Walker, H.J., Schwager, P., and Gilstrap, J.B. (2019). Is what you see what you get? Investigating the relationship between social media content and counterproductive work behaviors, alcohol consumption, and episodic heavy drinking. *The International Journal of Human Resource Management*, 30(15), 2251-2272.
- [12] Ben-Porath, Y.S., Corey, D.M., and Tarescavage, A.M. (2017). Using the MMPI-2-RF in preemployment evaluations of police officer candidates. In C.L. Mitchell and E.H. Dorian (Eds.), *Police Psychology and Its Growing Impact on Modern Law Enforcement*, 51-78. Hershey, PA: IQI Global.
- [13] Berkelaar, B.L. (2010). *Cyber-vetting: Exploring the implications of online information for career capital and human capital decisions*. Doctoral dissertation, Purdue University. Retrieved from <https://docs.lib.purdue.edu/dissertations/AAI3444477/>
- [14] Berkelaar, B.L. (2014). Cybervetting, online information, and personnel selection: New transparency expectations and the emergence of a digital social contract. *Management Communication Quarterly*, 28(4), 479-506.
- [15] Boes, J., Chandler, C.J., and Timm, H.W. (1997). *Police integrity: Use of personality measures to identify corruption-prone officers*. Technical Report PERS-TR-97-003. Monterey, CA: Defense Personnel Security Research Center.
- [16] Brown, V.R., and Vaughn, E.D. (2011). The writing on the (Facebook) wall: The use of social networking sites in hiring decisions. *Journal of Business and Psychology*, 26(2), 219.
- [17] Borofsky, G.L. (2017). User's manual for the Employee Reliability Inventory. Revision 6.5 CDN – 10/2004. Retrieved from www.psychometrics.com/wp-content/uploads/2018/08/ERI-Users-Manual-2.pdf
- [18] Borofsky, G.L. (2000). Predicting involuntary dismissal for unauthorized absence, lateness, and poor performance in the selection of unskilled and semiskilled British contract factory operatives: The contribution of the Employee Reliability Inventory. *Psychological Reports*, 87, 95-104.
- [19] Brodsky, S.L. (1978). [Test review of the Reid Report.] In O.K. Buros (Ed.). *The eighth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [20] Buss, D.M. (1987). Selection, evocation, and manipulation. *Journal of Personality and Social Psychology*, 53, 1214-1221.

- [21] Butcher, J., Derksen, J., Sloore, H., and Sirigatti, S. (2003). Objective personality assessment of people in diverse cultures: European adaptations of the MMPI-2, *Behaviour Research and Therapy*, 41(7), 819-840. doi.org/10.1016/S0005-7967(02)00186-9.
- [22] Butcher, J.N., Graham, J.R., Ben-Porath, Y.S., Tellegen, A., Dahlstrom, W.G., and Kaemmer, B. (2001). *MMPI-2 (Minnesota Multiphasic Personality Inventory-2) manual for administration, scoring, and interpretation (Rev. Ed.)*. Minneapolis, MN: University of Minnesota Press.
- [23] Clark, L.A., and Roberts, S.J. (2010). Employer's use of social networking sites: A socially irresponsible practice. *Journal of Business Ethics*, 95(4), 507-525.
- [24] Connolly, J. J., Kavanagh, E. J., and Viswesvaran, C. (2007). The convergent validity between self and observer ratings of personality: A meta-analytic review. *International Journal of Selection and Assessment*, 15(1), 110-117. doi.org/10.1111/j.1468-2389.2007.00371.x.
- [25] Creative Organizational Design (2019). The Reid Report. Retrieved from <https://www.creativeorgdesign.com/tests/reid-report-the/>
- [26] Crisi, A., and Palm, J. (2018). *The Crisi Wartegg System (CWS). Manual for administration, scoring and interpretation*. N.Y., USA: Routledge
- [27] Cunningham, M. R., Trucott, M., and Wong, D.T. (1990). An experimental investigation of integrity: Testing the predictive validity of the Reid Report. Unpublished manuscript.
- [28] Dages, K.D., Zimmer, S., and Jones, J.W. (2017). Pre-employment risk screening: Comparability of integrity assessment technology platforms. *International Journal of Selection and Assessment*, 25, 390-400. doi: 10.1111/ijsa.12193.
- [29] Dahling, J.J., Whitaker, B.G., and Levy, P.E. (2009). Development and validation of a new Machiavellianism Scale. *Journal of Management*, 35(2), 219-257.
- [30] Davison, H.K., Maraist, C., and Bing, M.N. (2011). Friend or foe? The promise and pitfalls of using social networking sites for HR decisions. *Journal of Business and Psychology*, 26(2), 153-159.
- [31] Davison, H.K., Maraist, C., Hamilton, R.H., and Bing, M.N. (2012). To screen or not to screen? Using the Internet for selection decisions. *Employee Responsibilities and Rights Journal*, 24(1), 1-21.
- [32] Detrick, P. and Chibnall, J.T. (2002). Prediction of police officer performance with the Inwald Personality Inventory, *Journal of Police and Criminal Psychology*, 17, 9-17.
- [33] DNV (2019). Testsertifisering [Test Certification]. Norway; Hövik: DNV (Det Norske Veritas). Retrieved 2 Mar 2, 2021, from https://www.dnvgl.no/services/testsertifisering-52540?_ga=2.203812307.2102773268.1615544643-1871511479.1615544643&_gac=1.247202480.1615544643.EAIaIQobChMIkNf14MSq7wIVcwzmCh2NYg4HEAAYASAAEgI8lvD_BwE
- [34] Domino, G. (1995). [Test review of the Reid Report]. In J.C. Conoley and J.C. Impara (Eds.). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

- [35] EFPA (2013). EFPA review model for the description and evaluation of psychological and educational tests. Retrieved from <http://www.efpa.eu/professional-development/assessment>
- [36] Funder, D.C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, 4, 652-670.
- [37] Funka (2021). Accessibility analyses. Retrieved March 12, 2021, from <https://www.funka.com/en/we-offer/accessibility-audits/>
- [38] Ganguli, H.C. (1985). [Test review of The Stanton Survey and the Stanton Survey Phase II]. In J.V. Mitchell, Jr. (Ed.). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [39] Gomez, A. (2019). Impact of conditional job offer on applicant reactions to social media in the selection process. (Publication No. 789) [Master's thesis, California State University San Bernardino]. *Electronic Theses, Projects, and Dissertations*.
- [40] Gueutal, H.G., Kluemper, D.H., and Rosen, P.A. (2009). Future employment selection methods: evaluating social networking web sites. *Journal of Managerial Psychology*, 24(6), 567-580
- [41] Guion, R.M. (1995). [Test review of Employee Reliability Inventory]. In J.C. Conoley and J.C. Impara (Eds.). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [42] Handler, M., Honts, C.R., Krapohl, D.J., Nelson, R. and Griffin, S. (2009). Integration of pre-employment polygraph screening into the police selection process. *Journal of Police and Criminal Psychology*, 24, 69-86. doi.org.10.1007/s11896-009-9050-2.
- [43] Harris, W.G. (1985). [Test review of The Stanton Survey and the Stanton Survey Phase II]. In J.V. Mitchell, Jr. (Ed.). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [44] He, Y., and Gong, R. (2010). A review of integrity tests in personnel selection. In 2010 International Conference on E-Product E-Service and E-Entertainment. 1-6.
- [45] Hogan, R. (2009). *The Hogan Development Survey technical manual*. Tulsa, OK, USA: Hogan Assessment Systems.
- [46] Hollwitz, J.C., and Pawlowski, D.R. (1997). The development of a structured ethical integrity interview for pre-employment screening. *The Journal of Business Communication*, 34(2), 203-219.
- [47] Institute for Personality and Ability Testing, Inc. (2012). *Inwald Personality Inventory-2*. Retrieved from https://post.ca.gov/portals/0/post_docs/publications/psychological-screening-manual/IPI-2.pdf
- [48] Jeanrie, C. (1998). [Test review of PSC-Survey ADT]. In J.C. Impara and B.S. Blake (Eds.). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [49] Jones, J.W., Brasher, E.E., and Huff, J.W. (2002). Innovations in integrity-based personnel selection: Building a technology-friendly assessment. *International Journal of Selection and Assessment*, 10(1-2), 87-97.

- [50] Jones, J.W., and Dages, K.D. (2003). Technology trends in staffing and assessment: A Practice note. *International Journal of Selection and Assessment*, 11(2/3), 247-252
- [51] Krapohl, D.J. (2002). The polygraph in personnel screening. In M. Kleiner (Ed.), *Handbook of polygraph testing*, 217-236. Academic Press, San Diego, CA
- [52] Kluemper, D.H., and Rosen, P.A. (2009). Future employment selection methods: evaluating social networking web sites. *Journal of Managerial Psychology*, 24(6), 567-580.
- [53] Kluemper, D.H., Rosen, P.A., and Mossholder, K.W. (2012). Social networking websites, personality ratings, and the organizational context: More than meets the eye? *Journal of Applied Social Psychology*, 42(5), 1143-1172.
- [54] Liberman, B., Seidman, G., Mckenna, K.Y., and Buffardi, L.E. (2011). Employee job attitudes and organizational characteristics as predictors of cyberloafing. *Computers in Human Behavior*, 27(6), 2192-2199.
- [55] Madera, J.M. (2012). Using social networking websites as a selection tool: The role of selection process fairness and job pursuit intentions. *International Journal of Hospitality Management*, 31(4), 1276-1282.
- [56] McCoy-Arballo, T.J. (2015). A correlational study of the empirical scales on the M-PULSE Inventory and the Inwald Personality Inventory-2 (UMI 3685777) [Doctoral dissertation, Alliant International University]. ProQuest Dissertation and Theses Global.
- [57] Murphy, K.R. (1995). [Test review of the Reid Report]. In J.C. Conoley and J.C. Impara (Eds.). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [58] National Research Council (2003). *The polygraph and lie detection. Committee to review the scientific evidence on the polygraph. Division of Behavioral and Social Sciences and Education. The National Academies Press, Washington, DC.*
- [59] Newness, K., Steinert, J., and Viswesvaran, C. (2012). Effects of personality on social network disclosure: Do emotionally intelligent individuals post inappropriate content? *Psihologijske Teme*, 21(3), 473-486.
- [60] Office of Technology Assessment (1983). *Scientific validity of polygraph testing: A research review and evaluation*. Congress of the United States, Office of Technology Assessment.
- [61] Ones, D.S. (1993). *The construct validity of integrity tests*. Unpublished doctoral dissertation. Iowa: University of Iowa City.
- [62] Ones, D.S., and Viswesvaran, C. (1998). The effects of social desirability and faking on personality and integrity assessment for personnel selection. *Human Performance*, 11(2-3), 245-269. doi.org/10.1080/08959285.1998.9668033.
- [63] Ones, D.S., Viswesvaran, C., and Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679-703. DOI: 10.1037/0021-9010.78.4.679.

- [64] Pearson, L.C. (1998). [Test review of PSC-Survey-SA]. In J.C. Impara and B.S. Blake (Eds.). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [65] Psykologisk Metod AB (2020). UPP™ – Understanding Personal Potential samt UPP StepOne Screen™. *Användarhandbok, Åttonde upplagan* [User Manual. 8th ed.]. Psykologisk Metod AB.
- [66] Psykologisk Metod AB (2019). UPP™ – Vetenskaplig bakgrund, elfte upplagan. [Scientific background, 11th ed.]. Psykologisk Metod AB.
- [67] Psykologisk Metod AB (2011). UPP-testet version 2.0 Understanding Personal Potential. *Teknisk manual* [Technical Manual]. Danderyd, Sweden: Psykologisk Metod AB. Retrieved March 15, 2021, from <https://docplayer.se/7032632-Upp-testet-version-2-0-understanding-personal-potential-teknisk-manual.html>
- [68] Raskin, R., and Terry, H. (1988). A principal-component analysis of the narcissistic personality inventory and further evidence of its construct validity. *Journal of Personality and Social Psychology*, 54(5), 890-902.
- [69] Reid Report, The. *Creative Organizational Design*. Retrieved from <https://www.creativeorgdesign.com/tests/reid-report-the/>
- [70] Roberts, R.M., Tarescavage, A.M., Ben-Portath, Y.S., and Roberts, M.D. (2019). Predicting postprobationary job performance of police officers using CPI and MMPI-2-RF test data obtained during preemployment psychological screening. *Journal of Personality Assessment*, 101(5), 44-555. doi: doi.org/10.1080/00223891.2018.1423990.
- [71] Rosengren, J., and Svanå, S. (2017). Like it or not-the recruitment practises of today a qualitative study on how social network sites are used within recruitment. (Publication No. 138) [Master's thesis, University of Gothenburg]. Gothenburg University Publications Electronic Archive.
- [72] Rudner, L.M. (1995). [Test review of Employee Reliability Inventory]. In J.C. Conoley and J.C. Impara (Eds.). *The twelfth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [73] Sackett, P.R. (1994). Integrity testing for personnel selection. *Current Directions in Psychological Science*, 3(3), 73-77.
- [74] Sackett, P.R., and Wanek, J.E. (1997). Integrity testing: An overview. *Security Journal*, 8, 11-18.
- [75] Sackett, P.R., and Wanek, J.E. (1996). New developments in the use of measures of honesty integrity, conscientiousness, dependability trustworthiness, and reliability for personnel selection. *Personnel Psychology*, 49(4), 787-829.
- [76] Schmidt, F.L., and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262.
- [77] Schneider, T.J. (2015). *Social networking sites and personnel Selection: An initial validity assessment*. (Publication No. 3367) [Doctoral dissertation, Western University]. Electronic Thesis and Dissertation Repository.

- [78] Sellbom, M., Fischler, G.L., and Ben-Porath, Y.S. (2007). Identifying MMPI-2 predictors of police officer integrity and misconduct. *Criminal Justice and Behavior*, 34(8), 985-1004. doi:10.1177/0093854807301224
- [79] Sjøberg, S, and Sjøberg, A. (2017). MINT Technical Manual. 2017 Edition. Stockholm, Sweden: Assessio International AB. Retrieved May 25, 2021, from https://support.assessio.com/hc/en-us/article_attachments/360008783598/MINT-technical-manual-2017.pdf
- [80] Sjoeborg, L. (2015). Correction for faking in self-report personality tests. *Scandinavian Journal of Psychology*, 56, 582-591. doi: 10.1111/sjop.12231.
- [81] Sjoeborg, L. (2008a). Bortom Big Five: Konstruktion och validering av ett personlighetstest [Beyond the Big Five: Construction and validation of a personality test].
- [82] Sjoeborg, L. (2008b). Emotional intelligence and life adjustment. In J.C. Cassady and M.A. Eissa (Eds.), *Emotional Intelligence: Perspectives on Educational & Positive Psychology*, 169-183. New York: Peter Lang Publishing. Retrieved 19 May, 2021, from <https://core.ac.uk/download/pdf/7281335.pdf>
- [83] Sjoeborg, S., Svensson, C., and Sjoeborg, A. (2019). *Measuring and Assessing Individual Potential – Technical Manual*. Stockholm, Sweden: Assessio International AB. Retrieved March 15, 2021, from https://support.assessio.com/hc/en-us/article_attachments/4404011603857/MAP_TechManual_20200625__3__1_.pdf
- [84] Skoglund, T.H. (2018a). Egnert som politi? Seleksjonsmetodenes prediktive validitet ved PHS sin egnethetsvurdering [The Predictive Validity of the Police Student Selection in Norway]. Retrieved from <https://phs.brage.unit.no/phs-xmlui/handle/11250/2501364>
- [85] Skoglund, T.H. (2018b). Integrity testing in the Norwegian Police selection. Presentation at the 60th International Military Testing Association (IMTA) Conference, Kingston, Canada.
- [86] Society for Human Resources Management (2008). Online technologies and their impact on recruitment strategies. Retrieved from <https://www.shrm.org/search/pages/default.aspx#k=Online%20technologies%20and%20their%20impact%20on%20recruitment%20strategies.&filters=>
- [87] Society for Industrial and Organizational Psychology (2018). *Principles for the validation and use of personnel selection procedures* (5th ed.). Bowling Green, Ohio: Society for Industrial and Organizational Psychology.
- [88] Stoughton, J.W. (2011). Examining applicant reactions to the use of social networking websites in pre-employment screening [Master's thesis, North Carolina State University]. North Carolina State University Library Repository.
- [89] Stoughton, J.W., Thompson, L.F., and Meade, A.W. (2013). Big five personality traits reflected in job applicants' social media postings. *Cyberpsychology, Behavior, and Social Networking*, 16(11), 800-805.
- [90] Stoughton, J.W., Thompson, L.F., and Meade, A.W. (2015). Examining applicant reactions to the use of social networking websites in pre-employment screening. *Journal of Business and Psychology*, 30(1), 73-88.

- [91] Sundre, D.L. (1995). [Test review of Personnel Reaction Blank]. In J.C. Impara and B.S. Blake (Eds.). *The thirteenth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- [92] Tarescavage, A.M., Corey, D.M., and Ben-Porath, Y.S. (2016). A prorating method for estimating MMPI-2-RF scores from MMPI responses: Examination of score fidelity and illustration of empirical utility in the PERSEREC police integrity study sample. *Assessment*, 23(2), 173-190. doi: 10.1177/1073191115575070.
- [93] Tarescavage, A.M., Corey, D.M., and Ben-Porath, Y.S. (2015). Minnesota Multiphasic Personality Inventory–2 Restructured Form (MMPI–2–RF) predictors of police officer problem behavior. *Assessment*, 22, 116-132. doi:10.1177/1073191114534885.
- [94] Tippins, N.T. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior*, 2(1), 551-582.
- [95] Van Hein, J.L., Kramer, J.J., and Hein, M. (2007). The validity of the Reid Report for selection of corrections staff. *Public Personnel Management*, 36(3), 269-280.
- [96] Van Iddekinge, C.H., Lanivich, S.E., Roth, P.L., and Junco, E. (2016). Social media for selection? Validity and adverse impact potential of a Facebook-based assessment. *Journal of Management*, 42(7), 1811-1835.
- [97] Van Iddekinge, C.H., Roth, P.L., Raymark, P.H., and Odle-Dusseau, H.N. (2012). The criterion-related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499-530. doi.org/10.1037/a0021196.
- [98] Vidotto, G., Anselmi, P., Filipponi, L., Tommasi, M., and Saggino, A. (2018). Using overt and covert items in self-report personality tests: Susceptibility to faking and identifiability of possible fakers. *Frontiers in Psychology* [online]. doi.org/10.3389/fpsyg.2018.01100.
- [99] Viswesvaran, C. and Ones, D.S. (1997). A review of the Personnel Reaction Blank (PRB). *Security Journal*, 8, 129-132.
- [100] Viswesvaran, C. and Ones, D.S. (1997). Review of the Stanton Survey. *Security Journal*, 8, 167-169.
- [101] Wanek, J.E., Sackett, P.R., and Ones, D.S. (2003). Towards an understanding of integrity test similarities and differences: An item-level analysis of seven tests. *Personnel Psychology*, 56(4), 873-894. doi.org/10.1111/j.1744-6570.2003.tb00243.x.
- [102] Westlin, E. (8 Apr, 2021). UPP Integrity Scale (tentative description, under development, based on Lennart Sjöberg). *Psykologisk Metod AB*.
- [103] Westlin, E. (2019). Validation study of specific narrow UPP scales to specific narrow job performance criteria (Internal paper). *Psykologisk Metod AB*.
- [104] Wheeler, K.G. (1985). [Test review of The Stanton Survey and the Stanton Survey Phase II]. In J.V. Mitchell, Jr. (Ed.). *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.

- [105] Witta, E.L. (1998). [Test review of PSC-Survey-SA]. In J.C. Impara and B.S. Blake (Eds.). The thirteenth mental measurements yearbook. Lincoln, NE: Buros Institute of Mental Measurements.
- [106] Zeidner, R. (2007). How deep can you probe? HR Magazine 52(10), 57-60.



Chapter 5 – DEFINING MORAL INTEGRITY IN THE MILITARY¹

Wendy Darr

Department of National Defence
CANADA

Markéta Zajícová

Czech Armed Forces
CZECH REPUBLIC

Kai Nyman

Finnish Defence Research Agency
FINLAND

Maria Fors Brandebo

Swedish Defence University
SWEDEN

Marjolein Verboom

Netherlands Armed Forces
NETHERLANDS

Gerhard Wolgers

Swedish Armed Forces
SWEDEN

Personnel selection research has witnessed a growing interest in integrity tests. Whether overt (e.g., counter-productive measures) or covert (e.g., personality-based) in nature, integrity tests have been found to predict workplace criteria from Counterproductive Work Behaviour (CWB) to job performance and absenteeism (e.g., Ref. [59]). However, these tests have faced long-standing criticism for their criterion-focussed nature [32]. Their content is largely dictated by the criteria that they intend to predict. Consequently, existing integrity tests appear to assess multiple constructs thought to predict CWBs, like dependability, honesty, and beliefs about whether CWBs should be punished [45], [60]. Responding to an American Psychological Association review of integrity tests, Ones [44] attempted to shed light on these issues by strongly recommending that researchers dedicate increased attention to “construct-oriented” matters (p. 18). In trying to answer the question, “what do integrity tests measure?” Ones [44] also conducted a factor analysis of all integrity tests to find evidence for a general Conscientiousness factor. However, this personality factor is a distinct construct of its own, and has also been shown to predict similar criteria [4].

A construct-oriented approach to test development involves the identification of important elements of integrity and developing content to reflect those elements [32]. But, what is integrity? Management and leadership researchers have asked this question many times. Palanski and Yammarino [48] highlighted the confusion and disagreement about the construct of integrity, and proposed a conceptual framework to guide its measurement. In a 2013 article, Simons, Palanski and Treviño shed further light on the issues surrounding integrity, particularly around its definition, measurement, antecedents, and outcomes. Simons et al. [54] conclude that while the concept of integrity has strong practical significance, it remains an underexplored area. As a result, clarification is required about its theoretical underpinnings, measurement, and nomological relationships with other constructs.

In general, personnel selection research has dedicated little attention to addressing the construct-related issues that have been highlighted by management/leadership researchers. The objective of this paper is to further an understanding of the integrity construct with the aim of answering the critical question, what should integrity tests measure?

¹ A revised version of this chapter has been published in *Military Psychology*.

5.1 EARLIER CLARIFICATIONS ON INTEGRITY

There are at least two papers in the leadership and management literature that have attempted to provide an understanding of integrity. These include Palanski and Yammarino's [48] and Bauman's [6] review of conceptual definitions and meanings of integrity. These publications are reviewed here with the intent of comparing and/or integrating these with other conceptions of integrity.

Palanski and Yammarino [48] divided integrity definitions into five main categories. The first refers to integrity in terms of wholeness and suggests a multi-faceted characterisation (in terms of behaviours, thoughts, emotions) of a person across situations and time (e.g., [64]). The second category refers to integrity as consistency between actions and words. It is assumed to be stable across time and situations (e.g., Ref. [53]). Their third category also views integrity in terms of word-action consistency, but particularly in adverse or challenging situations, or in the face of temptation when there is a choice between two courses of action (e.g., Ref. [47]). The fourth category regards integrity as being true to oneself, such that there is alignment between a person's words/actions and internalised values (e.g., Ref. [29]). The final category sees integrity in relation to morality (e.g., Ref. [5]).

To address the varied references to integrity, Palanski and Yammarino [48] propose the use of a virtue-based approach, defining virtues as components of good character or dispositions to behave in a certain way [37]. Drawing upon Audi and Murphy's [2] work, they distinguish between substantive and adjunctive virtues. Substantive virtues are morally good in nature. For instance, honesty and fairness are cited as substantive virtues. Adjunctive virtues are described as being neither good nor bad, but necessary for demonstrating substantive virtues. They characterise courage as adjunctive because one could exhibit the courage to harm (immoral behaviour) or the courage to defend others from harmful behaviour (moral behaviour). Using this approach, Palanski and Yammarino [48] argue that three categories of integrity definitions (i.e., word-action consistency, word-action consistency in adversity, true to oneself) are based on adjunctive virtues like authenticity and courage. Only the morally based definition is seen as based on substantive virtues, including honesty, trustworthiness, fairness, and compassion. Finally, the definition that reflects on integrity in terms of wholeness is regarded as comprising both sets of virtues.

Bauman [6] similarly attempts to clarify the concept of leadership integrity by conducting a review of definitions in the business literature. He acknowledges Palanski and Yammarino's [48] non-moral and moral definitions, but argues for a moral foundation by reviewing integrity's historical and philosophical basis. He describes early Roman governance as referring to integrity (among other values) in a call for ethical leadership. He also cites dictionary definitions of integrity where it is described in terms of moral soundness (e.g., the 1913 edition of the Webster's dictionary). Bauman [6] then reviews three key leadership theories, namely: transformational [11], authentic [22], and spiritual [21]. Each demonstrates the relevance of moral values, beliefs, and principles. Bauman then proposes that leaders with moral integrity act in accordance with principles or values, and that this behaviour occurs because of the leader's identity-conferring commitment to moral values.

5.2 ADVANCING AN UNDERSTANDING OF INTEGRITY

In this section, we integrate these earlier reviews on leadership integrity with philosophical accounts to advance a deeper understanding of the integrity construct.

5.2.1 Nature of Values and Integrity

Palanski and Yammarino [48] and Bauman [6] use different approaches to conceptualise integrity, but both agree that it acts or behaves in accordance with values or principles. But they differ in their understanding of the

nature of these same values/principles. Palanski and Yammarino [48] cite values/principles commonly highlighted in the literature (e.g., honesty, justice/fairness, trustworthy, and compassion). Meanwhile, Bauman [6] mentions only four (e.g., honesty, keeping promises, fairness, not harming innocent people). From the perspective of measurement, these divergent lists pose a problem. It is not clear whether one should focus on the common values from both, or if measuring all of them would sufficiently capture integrity.

Advancing an understanding of integrity and its correlates in organisational research presents another challenge. In Mayer, Davis, and Schoorman's [40] integrative model of organisational trust, integrity is seen as an antecedent of trust. Consequently, managers who are perceived by employees as demonstrating integrity are more likely to be trusted. But, if integrity is also defined and measured in terms of trust, overlapping construct variance will not provide a fair test of this model. Further problems are evident in Colquitt, Scott, and LePine's [13] meta-analysis of trust and its correlates, one of which is integrity. In their examination, integrity is defined using Mayer et al.'s [40] definition as perceived adherence to a set of principles that the perceiver finds acceptable. It is further operationalised by coding measures of credibility, promise keeping, and procedural justice as integrity. Clearly, there is a need to identify specific values and principles that are most pertinent to integrity. To do this, we will briefly review philosophical accounts of integrity.

5.2.2 Philosophical Views on Integrity

The identification of specific values, commitments, or principles has been the primary source of integrity's definitional inconsistencies. This is reflected in the Stanford Encyclopedia of Philosophy's (SEP; [14]) summary of the five main philosophical accounts of integrity. These are briefly reviewed here to underscore the importance of specifying the values/commitments/principles pertinent to integrity. The first account views integrity in terms of self-integration. In other words, individuals with integrity are thought to act in a manner that is consistent with their own desires and commitment (e.g., Ref. [20]). This view acknowledges that these desires may fluctuate and even conflict, but the self-integrated individual is able to successfully resolve this. The main criticism of this view is that although the individual may successfully address their competing desires/commitments, there are no clear external standards for judging the relevance or significance of these desires/commitments.

The second account of integrity is the identity view of integrity. It is similar to the first, in that individuals with integrity are generally thought to stay true to their commitments. The term commitment is used broadly to include an individual's deeply rooted intentions or promises (e.g., Ref. [63]). Once more, the main criticism is that there is no specification as to the nature of those commitments, some of which may be superficial or could be tied to moral and immoral institutions, causes, or people.

The self-constitution view of integrity is predicated on the view that individuals with integrity act on principles that would be rationally endorsed by one's future self (e.g., Ref. [33]). This view is concerned with coherence between a person's past and present life. While this account of integrity attempts to address the criticisms of the preceding two views by identifying a standard for evaluating the principles by which one acts, standards are still specific to the individual.

The fourth account sees integrity simply as standing up for something. It is the first account that considers an individual in relation to others. Individuals with integrity are believed to act in a manner that matters to others in their community (e.g., Ref. [12]). It assumes that individuals will act on their best judgement, and that this is formed by deliberating with others on what is valuable and worth pursuing in life. While this view attempts to constrain the kinds of commitments or values that should be pursued by an individual of integrity by referring to those of importance to a community. But it leaves the possibility open that such a community may constitute one dedicated to the pursuit of illegal or immoral action.

The fifth and final account sees integrity in terms of moral purpose. It constrains commitments to moral ones; individuals with integrity are thought to act with moral purpose, which includes pursuing commitments that have been considered or deliberated upon from a moral point of view (e.g., Ref. [28]). However, this view is also not without criticism, because it does not define moral purpose. This leaves it up to the individual’s idea of morality, which may be influenced by society’s conception.

5.2.3 Categorising Integrity Definitions

The varying accounts and definitions of integrity (e.g., Refs. [6], [14], [48]) make it clear that the central issue with existing notions and definitions concerns the kinds of commitments, values, and/or principles (collectively referred to as standards from here on) to which an individual’s actions/behaviours must align. With this in mind, we propose a framework (see Table 5-1) for organising standards in terms of their nature (moral or other) and their referent (internal to self or individual) or external to self or individual). The philosophical accounts and leadership integrity definitions reviewed in the preceding section are categorised in the figure according to the nature of the standards and their referent.

Table 5-1: A Categorisation of Existing Integrity Definitions/Perspectives by Referent and Nature of Values.

		Referent		
		Internal		External
Nature	Other	self-integration (Frankfurt, 1971); identity view (Williams, 1981); self-constitution (Korsgaard, 2009); behavioural integrity (Simons, 2002);	adjunctive virtues (Palanski and Yammarino, 2007)	standing up for something (Calhoun, 1995)
	Moral	moral purpose (Halfon, 1989) moral identity-conferring (Bauman, 2013)	substantive virtues (Palanski and Yammarino, 2007)	<i>proposed definition of moral integrity</i>

As shown in Table 5-1, Bauman’s [6] moral identity-conferring and Palanski and Yammarino’s [48] virtue accounts are placed in the middle of the table in terms of their referent. While they mention specific standards (e.g., honesty, justice/fairness, keeping promises, fairness), which suggests that these are not left up to an individual to determine (i.e., no internal referent), they do not specify any particular external referent. The figure reveals an empty block at the bottom right (i.e., standards that are of moral nature with external referent). This is

where a definition of moral integrity should reside. Therefore, we propose definitions (and measures) of moral integrity employed for use in organisational or applied settings must refer to values, commitments, or principles (i.e., standards) that are *moral* in nature and that have an *external* referent (i.e., determined to be of significance in the organisation or applied setting to which the individual belongs). In other words, a determination of an individual's integrity is contingent upon the nature and referent of the standards (i.e., moral-external) used to evaluate integrity-relevant behaviour.

We argue that a definition of integrity must include standards that are moral in nature and have an external referent (i.e., determined by some community or organisation rather than oneself or an individual). In the following paragraphs, we further explain the need for specifying moral standards and stipulating an external referent in defining integrity.

5.2.3.1 Moral Nature of Standards

Morality is concerned “with the distinction between right and wrong” [46]. This is essential to integrity, because deficient moral reasoning has been consistently implicated and empirically linked as an antecedent to delinquent or deviant behaviour (e.g., Ref. [19]). O’Boyle, Forsyth, and O’Boyle [43] also proposed moral judgement and reasoning as antecedents that influence one’s affective and cognitive processes. This in turn influences CWB. Examining workplace aggression, Spector, Fox, and Domagalski [56] suggest that anger interferes with moral reasoning, a higher-level cognitive process. Strong negative associations have also been reported between various forms of destructive leadership and ethical leadership. For example, de Hoogh and Hartog [17] reported a correlation of $-.54$ between despotic leadership and ethical leadership, referred to as morality or fairness leadership in their study. The correlation between ethical leadership and abusive supervision was also high in Detert, Treviño, Burris, and Andiappan’s [18] examination ($r = -.51$), and in Brown, Treviño and Harrison’s [10] study ($r = -.61$). Ng and Feldman [42] synthesised the correlation between ethical and destructive leadership to be $-.49$ using meta-analysis. Together, these findings suggest the presence of a common substantive base (i.e., lack of morality) that underlies counterproductive behaviour. The requirement for the moral nature of standards is also consistent with Bauman’s [6] defence of the moral status of integrity, in which he illustrated linkages between integrity and various ethical leadership theories and constructs. Defined as such, the term *moral integrity* is more applicable.

5.2.3.2 External Referent

The requirement for an external referent of these standards is consistent with Babbitt’s [3] argument that personal integrity must be contextualised socially to have meaning. Indeed, such a context is critical to illustrate certain values and commitments. The SEP also acknowledges that little attention has been devoted to understanding social and political influences on personal integrity [14]. Simons et al. [54] also mention how little scholars have explored how situational factors impact leader integrity. In this case, they are concerned with moderating influences (e.g., incentives) on the demonstration of leader integrity.

The influence of context is also evident in Martin, Keating, Resick, Szabo, Kwan, and Peng’s [38] thematic analysis of leader integrity across six countries to find variations in the nature of values used to describe leader integrity. For example, honesty was identified in 81.5% of leader integrity descriptions provided by managers in Ireland, but was relatively lower in descriptions from other countries (55.6% [Germany], 36% [Austria], 27.5% [China]). Sense of responsibility towards others was evident in 52% to 58% of the descriptions provided by managers from Germany, Austria, China, and Hong Kong, but completely absent from descriptions of leader integrity provided by managers from the United States (US) and Ireland. Fair and just was identified by 75% of managers from China as characteristic of leader integrity, but only by 28% of managers in the US. Variation in

the salience of values from one context to another or from one person to another is a limiting factor in the study of moral integrity. For example, Tomlinson, Lewicki, and Ash [57] found that employee organisational citizenship behaviours depended not only on the extent to which their supervisors exhibited behavioural integrity (i.e., word-action alignment), but also on the extent to which the employees' values were congruent with those of their supervisors. Tomlinson et al. [57] did not specify the nature of these values, which was acknowledged as a limitation of their study, but called for future research to clarify the content of such values.

5.3 DEFINING MORAL INTEGRITY IN THE MILITARY

We used the requirements specified above (i.e., moral nature and external referent of standards) to develop a definition of moral integrity for use in the military. Military organisations are distinguished from others by an ethos that inculcates a common understanding of what it means to serve [62]. Commonly referred to as the profession of arms or military professionalism, this understanding refers to a shared set of core beliefs and values that is expected to guide personnel in performing their duties [15], [58]. In the Royal Netherlands Armed Forces, military professionalism includes concepts of self-leadership, moral professionalism, and ownership. These are socialised with personnel via training and leadership programs [16]. Lindsay, Jackson, Watola, Foster, and Ramsey [36] discussed how threats to military professionalism can erode moral reasoning, and result in immoral or illegal actions. In other words, this shared understanding of common values and expectations among military personnel serves as a compass for discerning proper conduct. As such, it is aligned with the definition of morality offered earlier (i.e., distinction between right and wrong; [46]), and serves as a relevant referent for identifying common standards (i.e., values, principles, commitments) to which military personnel are expected to adhere.

5.3.1 Methodological Approach

Members and invited representatives of this RTG, which represented 13 different countries' armed forces, were asked, through a request for information, to provide scanned copies of *clearly stated high-level code/values of ethics/integrity/profession of arms* (e.g., information contained in official brochures/pamphlets that are well-known, promoted within the Armed Forces). If the original document(s) were not written in English, representatives were asked to provide an official translation in English. They were also asked to ensure translations were reviewed in order to ensure that the initial content was intact in meaning. The 13 nations that contributed codes/values of their nations' military included Australia, Belgium, Canada, Czech Republic, Denmark, Finland, United Kingdom (British Army), Germany, The Netherlands, New Zealand, Norway, Sweden, and the USA.

Except for two nations whose codes were analysed separately due to late receipt, an initial content analysis of core values/codes was undertaken by two members of the RTG with the aim of identifying how different armed forces defined integrity. Every document was analysed according to the first steps of the constant comparative method introduced by Glaser and Strauss [23]. This begins with the open coding process. The documents were examined line by line in order to identify descriptions related to integrity. Seven of the 13 nations (Australia, Belgium, Canada, United Kingdom (British Army), the Netherlands, New Zealand, United States) used the actual term integrity in their codes/values, and gave a more detailed description of it. Focussing only on these seven nations' descriptions of integrity using words resembling those in the written text from their original documents, codes were sorted into different categories. For example, two separate codes were identified from the Australian Defence Force values (e.g., "We behave honestly and ethically"), namely: 1) Behave honestly, and 2) Behave ethically. These were further sorted into two categories called 1) Honesty and truthfulness, and 2) High ethical standards.

Based on this initial review, a subsequent content analysis was conducted. This included a review of the core values/codes from all 13 nations (two nations codes were reviewed/coded using a slightly different process, wherein members submitted their input via email). Keeping in mind the definition of morality as being concerned with knowing what is right (versus wrong), five RTG members independently reviewed each nation's codes/core values to identify values or phrases reflecting proper personnel conduct or behaviour. In some nations' codes, the conditions surrounding the demonstration of a value was also specified and this was duly noted. Next, the authors shared their findings, to arrive at consensus on the inclusion of a particular value/phrase as reflective of personnel conduct or behaviour. In most cases, all five members agreed in their identification of the same value/phrase as reflections of right/wrong personnel conduct or behaviour. Only a few values/phrases had to be discussed further to arrive at consensus. A list of agreed upon values/phrases was then compiled for each nation, and grouped into categories based on similarity in meaning.

5.3.2 Findings

5.3.2.1 Initial Content Analysis

Findings from the initial content analysis, which focussed only on descriptions of the term integrity (used by six nations), resulted in the identification of eight different categories/aspects of integrity, presented below:

- Follow norms, rules, and regulations;
- High ethical standards (doing what is right);
- Fairness and justice;
- Maintain personal and the organisation's dignity and reputation;
- Do not take advantage of role/position;
- Responsibility;
- Honesty and truthfulness; and
- Build and maintain trust.

These categories reflect the differing values/principles evident in earlier reviews of integrity definitions. In addition, a few of the identified categories were, in some aspects, incompatible. For example, in a few cases integrity was defined (among other things) as "doing what is right" and in other cases as "following rules." Discussions around this led to the conclusion that "doing what is right" could imply simply "(not) breaking the rules." Looking at the six remaining nations, we also observed that several used the same type of descriptions identified above, but not in relation to the term integrity. Other related terms were used like courage, selflessness, and respect.

5.3.2.2 Subsequent Content Analysis

Findings from the second content analysis, which included a review of all core values/codes from all 13 nations, resulted in the identification of several common categories/themes (see Table 5-2). As shown in Table 5-2, values/phrases were grouped into categories that best reflected the content of codes. The two most common categories were *respecting others* and *honesty*. The codes presented in Table 5-2 reflect a set of standards (common across at least four nations) that guide military personnel behaviour and conduct. As such, we define moral integrity in the military as:

Acting in accordance with military and societal values, which includes respecting others, putting service before self, having courage, and being honest and accountable, particularly when there are no clear rules or when no one is monitoring.

Table 5-2: Results of Content Analysis of High-Level Codes from NATO Military Institutions.

Category	Sample Content from Codes	Frequency*
Acting in accordance with rules, regulations, and society values	Follow the rules; actions match mission; comply with regulations	8
Respecting others	Treated equal; supporting/helping others; considerate to others	10
Service before self	Interest in society over interest of self; selfless commitment	7
Courage	Speaking out against wrongdoing; be active, not passive	8
Honesty	Impartial advice; no mateship to cover up bad behaviour; truthful	9
Accountability	Accept personal responsibility for actions; using resources effectively and efficiently	6
Conditions	When there are no clear, specific rules; when no one is looking; even when it may be unpopular; when no one is watching	5

* Number of nations represented within a specific category

5.4 DISCUSSION

The objective of this chapter was to further understanding of the integrity construct, with the aim of guiding the measurement of integrity in military personnel selection. Drawing upon earlier efforts to understand integrity in the domains of leadership and philosophy, we propose that any definition of integrity must refer to actions or behaviours that are aligned with moral standards (i.e., concerned values, principles, expectations about the right way to behave or act), and that these standards must be prescribed by an external source of relevance or significance (in the present case, the military). Through a content analysis of documents containing codes/values pertaining to ethics/integrity/profession of arms that are well-known and promoted within 13 nations, we identified common moral standards that formed the basis of the proposed definition, which may be more appropriately referred to as moral integrity. While our proposed definition will likely echo the codes of conduct and ethics of any military’s profession of arms, the approach we used in defining integrity recognises the importance of context (i.e., external referent) in identifying moral standards. As context can influence the salience of values (e.g., Ref. [38]), each military may compare the moral standards or values included in this proposed definition against their respective codes of ethics, and adjust them, if required, to include ones that are more salient in their specific context.

5.4.1 Situational Aspects of the Definition

The context of actions/behaviours is an important element of the proposed definition. Moral integrity refers to acting in accordance with certain standards in situations when there are no clear rules about how to act and/or when the action/behaviour cannot be easily traced to the actor. This aspect of the proposed definition takes situational factors

into account that have traditionally been ignored. For example, Berry, Sackett, Weimann [9] reviewed developments in integrity test research to find sparse ties to situational factors, despite suggestions integrity test scores could possibly change when the situation is altered (e.g., Ref. [41]). In fact, situational factors in the larger domain of personnel selection tend to be seen as unwanted sources of variance [27]. This is because selection research is typically based upon the notion of behavioural consistency or generalisation across situations [51]. However, recent research suggests situational features of an assessment protocol cue applicants to provide certain behavioural responses of interest, allowing for better prediction of future behaviour in similar situations [30]. This research also found that individuals who were more accurate in their assessment of the situational demands of four assessment centre exercises performed better, and that their situational assessment was related to future performance (i.e., on-the-job performance). Inherent in our definition of moral integrity is the specification of contextual requirements (i.e., lack of rules or monitors) which has implications for the way this construct is measured.

5.4.2 Implications for Assessing Integrity

As defined, moral integrity pertains to actions or behaviours. In providing guidance on the assessment of these actions/behaviours, we draw upon Lievens and Sackett's [35] discussion on predictor methods and the modular approach to personnel selection. Predictor methods determine how a construct is assessed, and refers to techniques or approaches used to elicit and gather information on the construct [1], [35]. Predictor methods can vary in a number of ways (e.g., structured/unstructured, written/oral response), called predictor method factors. Lievens and Sackett [35] described seven such factors: stimulus format, contextualisation, stimulus presentation consistency, response format, response evaluation consistency, information source, and instructions. Of these, contextualisation deserves consideration in the assessment of moral integrity.

Contextualisation pertains to providing contextual details in assessing a construct. According to Johns [31], contextual details can refer to the omnibus context (e.g., general descriptions of when, where, who, or why), or they can be more discrete (e.g., descriptions about the specific task, social environment, or physical environment). As discussed in Lievens and Sackett [35], assessments can be decontextualised (e.g., personality statements about one's personal preference or tendency), or have low contextualisation (e.g., personality statements about personal preferences with a condition such as "at work" added to the statements); medium contextualisation (e.g., interview questions that elicit responses requiring a general consideration about when, where, why, and how certain actions/behaviours unfold); or high contextualisation (e.g., role play exercises that provide information about the specific task, players in the social environment, as well as the physical setting).

In their review of the impact of contextualisation on the validity of assessments, Lievens and Sackett [35] found assessments with more contextualisation had higher validities and lower error variance compared to those with no contextualisation. Given the proposed definition of moral integrity, we believe that assessment methods that allow for medium to high levels of contextualisation will allow for a more valid assessment. Such methods may include a structured interview, Situational Judgement Test (SJT), or role play, and each could differ in the nature of method factors used (e.g., the stimulus format could be textual, face-to-face, or pictorial).

The level of contextualisation will require some experimentation to ensure that it does not affect the validity and transportability of the assessment. For example, Lievens [34] compared features of traditional SJTs and newer construct-driven SJTs. While traditional SJTs have higher contextualisation, scores from such measures tend to assess multidimensional constructs (yielding low convergent validity), and are less easily generalisable across settings (e.g., cultures, jobs). Construct-driven SJTs, while lower in contextualisation, are unidimensional (yielding high convergent validity), but may lower criterion-related validity due to decreased relevance to the criterion. Consequently, the level of contextualisation must be carefully balanced to optimise validity and

transportability. In addition, any selection procedure content should be based on work-related behaviours or Knowledge, Skills, Abilities, and Other Requirements (KSAOs) expected of the prospective employee before training or selection into the job [55].

Guenole, Chernyshenko, and Weekly [26] provided guidance in enhancing the validity of SJTs, which is also useful to the development of other assessment tools (e.g., interviews). Of particular relevance to contextualisation, they suggested that the generation of scenarios or contextual information must be specific to the construct dimension, and is best done by subject matter experts based on realistic work-related situations. However, it is the role of the industrial/organisational psychologist to ensure that the situation/scenario elicits the traits or behaviours of interest [26]. In other words, the psychologist must pay attention to the psychologically important characteristics of the situation more than the objective details provided by SMEs (e.g., specifics about who, when, or where). The importance of subjective perception of situations is highlighted in Rauthmann et al.'s [50] taxonomy of psychologically important, meaningful and consequential situation characteristics. They identified eight characteristics, the situational DIAMONDS (Duty, Intellect, Adversity, Mating, Positivity, Negativity, Deception, and Sociality), that were found to consistently depict situations and elicit related behaviours. Another similar, but lexically-derived seven dimension taxonomy of psychological situations, has been developed by Parrigon, Woo, Tay, and Wang [49] and is called CAPTION (Complexity, Adversity, Positive Valence, Typicality, Importance, Humor, and Negative Valence).

These taxonomies are useful tools to use in the development of scenario-based assessments, and can be used to compare the similarity of original SME-generated scenarios and ones that have been adapted by the psychologist to improve transportability or applicability to pre-employment contexts (i.e., for use with applicants who have no prior job knowledge or experience). Situations that are psychologically similar, even though they differ in objective characteristics, should elicit similarly relevant behaviours [50], [52]. Applying this guidance to the proposed definition of moral integrity implies that even though scenarios are generated around a particular military value (e.g., putting service before self) or realistic work situation that reflects the conditional aspects of the definition (e.g., lack of clarity in military regulations/policy), adapting these for use in selection would require ensuring that the key psychological characteristic(s) of the situation are retained. For example, if an SME-generated situation around this particular value reflects Rauthmann et al.'s, [50] psychological characteristic of Duty (e.g., being counted upon to do something) or Adversity (e.g., another person is under threat), care must be taken to ensure that the adapted scenario reflects these same psychological characteristics even if the objective details of the scenario are changed to a non-work context (e.g., public social context, school/classroom context).

5.4.3 Implications for Research on Integrity

As mentioned earlier, existing research on integrity in the personnel selection domain has largely used covert (e.g., personality-based) or overt (e.g., past deviant behaviours such as theft, cheating) measures to predict workplace outcomes, particularly CWBs [59]. We consider our proposed definition of integrity, as a behaviour, to be more proximal to the prediction of CWBs. In the case of personality predictors, we propose that at a general level personality predisposes individuals to act in accordance with certain moral standards in unclear/unmonitored situations, which in turn reduces or lowers CWBs. In other words, moral integrity (as per our proposed definition) is likely to mediate the relationship between personality and CWBs.

This is aligned with Marcus and Schuler's [39] general perspective on counterproductive behaviour, which is rooted in Gottfredson and Hirschi's [25] theory of crime in which self-control is thought to be more proximal to criminal behaviour. Marcus and Schuler [39] found empirical support for the core influence of self-control in accounting for CWBs. In Gottfredson and Hirschi's [25] explanations of self-control and criminal behaviour, they emphasised that situational variables such as the opportunity to act in a deviant manner is a necessary

condition for crime. Marcus and Schuler [39] empirically measured organisational monitoring to find that it was negatively related to CWBs. Therefore, we believe that our proposed definition of moral integrity is aligned with theoretical underpinnings of CWBs, and has the potential to advance integrity research when measured as defined. We encourage the development of integrity measures for research purposes that are aligned with our proposed definition, and the pursuit of related research.

5.5 CONCLUDING REMARKS

In closing, we believe that the contributions of this chapter will advance an understanding and measurement of moral integrity in military settings and the larger field of psychology.

5.6 REFERENCES

- [1] Arthur, W., Jr., and Villado, A.J. (2008). The importance of distinguishing between constructs and methods when comparing predictors in personnel selection research and practice. *Journal of Applied Psychology*, 93, 435-442. [dx.doi.org/10.1037/0021-9010.93.2.435](https://doi.org/10.1037/0021-9010.93.2.435)
- [2] Audi, R., and Murphy, P.E. (2006). The many faces of integrity. *Business Ethics Quarterly*, 16, 3-21.
- [3] Babbitt, S.E. (1997). Personal integrity, politics and moral imagination. In S. Brennan, T. Isaacs, and M. Milde (Eds.), *A Question of Values: New Canadian Perspectives on Ethics and Political Philosophy*, 107-31. Atlanta: Rodopi.
- [4] Barrick, M.R. and Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-24.
- [5] Batson, C.D., Thompson, E.R., Seufferling, G., Whitney, H., and Strongman, J.A. (1999). Moral hypocrisy: Appearing moral to oneself without being so. *Journal of Personality and Social Psychology*, 77, 525-537.
- [6] Bauman, D.C. (2013). Leadership and the three facets of integrity. *The Leadership Quarterly*, 24, 414-426.
- [7] Becker, T.E. (1998). Integrity in organizations: Beyond honesty and conscientiousness. *Academy of Management Review*, 23(1), 154-161.
- [8] Becker, T.E. (2001). Development and validation of a situational judgement test of employee integrity. *International Journal of Selection and Assessment*, 13, 225-232.
- [9] Berry, C.M., Sackett, P.R., and Weimann, S. (2007). A review of recent developments in integrity test research. *Personnel Psychology*, 60, 271-301.
- [10] Brown, M.E., Treviño, L.K., and Harrison, D.A. (2005). Ethical leadership: A social learning perspective for construct development and testing. *Organizational Behavior and Human Decision Processes*, 97, 117-134.
- [11] Burns, J.M. (1978). *Leadership*. New York: Harper & Row.
- [12] Calhoun, C. (1995). Standing for something. *Journal of Philosophy*, XCII, 235-260.

- [13] Colquitt, J.A., Scott, B.A., and LePine, J.A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology*, 92(4), 909-927. doi.org/10.1037/0021-9010.92.4.909
- [14] Cox, D., La Caze, M., and Levine, M. (2017, Spring). Integrity. *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (Ed.), Retrieved from <https://plato.stanford.edu/archives/spr2017/entries/integrity/>
- [15] Department of National Defence (2003). *Duty with honour: The Profession of Arms in Canada*. Ottawa, ON: Canadian Defence Academy Press.
- [16] de Graaf, M.C., and van Gils, M.J. (2012). *Military professionalism: An organizational challenge by itself*. In J. Stouffer and D. Lindsay (Eds.), *Threats to military professionalism: International perspectives*. Kingston, ON: Canadian Defence Academy Press.
- [17] de Hoogh, A.H.B., and Hartog, D.N.D. (2008). Ethical and despotic leadership, relationships with leader's social responsibility, top management team effectiveness and subordinates' optimism: A multi-method study. *Leadership Quarterly*, 19, 297-311.
- [18] Detert, J.R., Treviño, L.K., Burris, E.R., and Andiappan, M. (2007). Managerial modes of influence and counterproductivity in organizations: A longitudinal business-unit-level investigation. *Journal of Applied Psychology*, 92(4), 993-1005.
- [19] Dilchert, D.S., Ones, R.D., Davis, C.D., and Rostow, C.D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology*, 92(3), 616-627.
- [20] Frankfurt, H. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, LXVIII, 5-20.
- [21] Fry, L.W. (2003). Toward a theory of spiritual leadership. *The Leadership Quarterly*, 14, 693-727.
- [22] Gardner, W.L., Avolio, B.J., Luthans, F., May, D.R., and Walumbwa, F. (2005). "Can you see the real me?" A self-based model of authentic leader and follower development. *The Leadership Quarterly*, 16, 243-272.
- [23] Glaser, B., and Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Mill Valley, CA: Sociology Press.
- [24] Goldberg, L.R., Grenier, J.R., Guion, R.M., Sechrest, L.B., and Wing, H. (1991). *Questionnaires used in the prediction of trustworthiness in pre-employment selection decisions: An A.P.A. task force report*. Washington DC: American Psychology Association.
- [25] Gottfredson, M.R., and Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- [26] Guenole, N., Chernyshenko, O.S., and Weekly, J. (2017). On designing construct driven situational judgment tests: Some preliminary recommendations. *International Journal of Testing*, 17, 234-252. doi:10.1080/15305058.2017.1297817

- [27] Guion, R.M. (1991). Personnel selection, assessment, and placement. In M.D. Dunnette and L.M. Hough (Eds.). *Handbook of industrial and organizational psychology* (Vol. 2). Palo Alto, CA: Consulting Psychologists Press.
- [28] Halfon, Mark, 1989. *Integrity: A philosophical inquiry*. Philadelphia: Temple University Press.
- [29] Howell, J.M. and Avolio, B.J. (1995). Charismatic leadership: Submission or liberation? *Business Quarterly*, 60, 62-71.
- [30] Jansen, A., Melchers, K.G., Lievens, F., Kleinmann, M., Brändli, M., Fraefel, L., and König, C.J. (2013). Situation assessment as an ignored factor in the behavioral consistency paradigm underlying the validity of personnel selection procedures. *Journal of Applied Psychology*, 98(2), 326. doi: 10.1037/a0031257.
- [31] Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*, 31, 386-408.
- [32] Karren, R.J., and Zacharias, L. (2007). Integrity tests: Critical issues. *Human Resource Management Review*, 17(2), 221-235. doi: 10.1016/j.hrmr.2007.03.007.
- [33] Korsgaard, C.M. (2009). *Self-Constitution: agency, identity, and integrity*. Oxford: Oxford University Press.
- [34] Lievens, F. (2017). Construct-driven SJTs: Towards an agenda for future research. *International Journal of Testing*, 17, 269-276.
- [35] Lievens, F., and Sackett, P.R. (2017). The effects of predictor method factors on selection outcomes: A modular approach to personnel selection procedures. *Journal of Applied Psychology*, 102(1), 43-66. doi.org/10.1037/apl0000160
- [36] Lindsay, D.R., Jackson, R.J., Watola, D.J., Foster, C.A., and Ramsey, R.T. (2012). Threats to professionalism in the United States Military. In J. Stouffer and D. Lindsay (Eds.). *Threats to military professionalism: International perspectives*. Kingston, ON: Canadian Defence Academy Press.
- [37] MacIntyre, A. (1984). *After virtue* (2nd ed.). Notre Dame, IN: University of Notre Dame Press.
- [38] Martin, G.S., Keating, M.A., Resick, C.J., Szabo, E., Kwan, H.K., and Peng, C. (2013). The meaning of leader integrity: A comparative study across Anglo, Asian, and Germanic cultures. *The Leadership Quarterly*, 24, 445-461.
- [39] Marcus, B., and Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology*, 89(4), 647-660. doi.org/10.1037/0021-9010.89.4.647.
- [40] Mayer, R.C., Davis, J.H., and Schoorman, F.D. (1995). An interactive model of organizational trust. *Academy of Management Review*, 20, 709-734.
- [41] Mumford, M.D., Connelly, M.S., Helton, W.B., Strange, J.M. and Osburn, H.K. (2001). On the construct validity of integrity tests: Individual and situational factors as predictors of test performance. *International Journal of Selection and Assessment*, 9, 240-257.

- [42] Ng, T.W.H., and Feldman, D.C. (2015). Ethical leadership: Meta-analytic evidence of criterion related and incremental validity. *Journal of Applied Psychology*, 100(3), 948-965. doi.org/10.1037/a0038246
- [43] O'Boyle, E.H., Jr., Forsyth, D.R., and O'Boyle, A.S. (2011). Bad apples or bad barrels: An examination of group and organizational level effects in the study of individual counterproductive work behavior. *Group and Organizational Management*, 36, 39-69.
- [44] Ones, D.S. (1993). The construct validity of integrity tests. Unpublished doctoral dissertation. Iowa: University of Iowa City.
- [45] Ones, D.S., Viswesvaran, C., and Schmidt, F.L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78(4), 679-703.
- [46] Oxford University Press (1976). *Concise Oxford dictionary*. Author.
- [47] Paine, L.S. (2005). Integrity. In P.H. Werhane and R.E. Freeman (Eds.), *The Blackwell encyclopedia of management: Business ethics* (2nd ed.). Malden: MA: Blackwell Publishing.
- [48] Palanski, M.E., and Yammarino, F.J. (2007). Integrity and leadership: Clearing the conceptual confusion. *European Management Journal*, 25(3), 171-184. doi:10.1016/j.emj.2007.04.006
- [49] Parrigon, S., Sang, S.E., Tay, L., and Wang, T. (2017). CAPTION-ing the situation: A lexically-derived taxonomy of psychological situation characteristics. *Journal of Personality and Social Psychology*, 112, 642-681.
- [50] Rauthmann, J.F., Gallardo-Pujol, D., Guillaume, E.M., Todd, E., Nave, C.S., Sherman, R.A., et al. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, 107, 677-718.
- [51] Schmitt, N., and Ostroff, C. (1986). Operationalizing the "behavioral consistency" approach: Selection test development based on a content-oriented strategy. *Personnel Psychology*, 39(1), 91-108.
- [52] Sherman, R.A., Nave, C.N., and Funder, D.C. (2010). Situational similarity and personality predict behavioral consistency. *Journal of Personality and Social Psychology*, 99, 330-343. doi:10.1037/a0019796
- [53] Simons, T. (2002). Behavioral integrity: The perceived alignment between managers' words and deeds as a research focus. *Organization Science*, 13, 18-35.
- [54] Simons, T.L., Palanski, M., and Trevino, L. (2013). Toward a broader – but still rigorous – definition of leader integrity: Commentary. *The Leadership Quarterly*, 24, 1-12. Retrieved from <http://scholarship.sha.cornell.edu/articles/725>
- [55] Society for Industrial and Organizational Psychology. (2018). *Principles for the validation and use of personnel selection procedures* (Fifth ed.) Society for Industrial and Organizational Psychology.
- [56] Spector, P.E., Fox, S., and Domagalski, T. (2005). Emotions, violence, and counterproductive work behavior. In E.K. Kelloway, J. Barling, and J. Hurrell (Eds.), *Handbook of Workplace Violence*, 29-46. Thousand Oaks, CA: Sage.

- [57] Tomlinson, E.C., Lewicki, R.J., and Ash, S.R. (2014). Disentangling the moral integrity construct: A values congruence as a moderator of the behavioral integrity-citizenship relationship. *Group and Organization*, 39(6), 720-743.
- [58] United States Government Accountability Office (2015). *Military personnel: Additional steps are needed to strengthen DOD's oversight of ethics and professionalism issues* (Report to Congressional Committees, GAO 15-711). Washington, D.C: Author
- [59] Van Iddekinge, C.H., Roth, P.L., Raymark, P.H., and Odle-Dusseau, H.N. (2012). The criterion related validity of integrity tests: An updated meta-analysis. *Journal of Applied Psychology*, 97(3), 499-530.
- [60] Van Iddekinge, C.H., Taylor, M.A., and Eidson, Jr., C.E. (2005). Broad versus narrow facets of integrity: Predictive validity and subgroup differences. *Human Performance*, 18(2), 151-177.
- [61] Vogelgesang, G.R., Leroy, H., and Avolio, B.J. (2013). The mediating effects of leader integrity with transparency in communication and work engagement/performance. *The Leadership Quarterly*, 24(3), 405-413. doi.org/10.1016/j.leaqua.2013.01.004.
- [62] Walzer, M. (1981). Two kinds of military responsibility. *Parameters: Journal of the US Army War College*, 11(1), 42-26.
- [63] Williams, B. (1981). *Moral Luck: Philosophical Papers 1973 – 1980*. Cambridge: Cambridge University Press. Doi: doi.org/10.1017/CBO9781139165860.004.
- [64] Worden, S. (2003). The role of integrity as a mediator in strategic leadership: A recipe for reputational capital. *Journal of Business Ethics*, 46, 31-44.



Chapter 6 – MILITARY PERSONNEL SELECTION AND TECHNOLOGY: OVERVIEW

Elliot Loh

Department of National Defence
CANADA

Annemie Defranc

Belgian Defence
BELGIUM

6.1 BACKGROUND

Military organisations must recruit large numbers of personnel to staff a diverse range of occupations. To this end, the militaries represented in the Research Task Group (RTG) have developed unique personnel assessment and selection processes to effectively review applicants and identify suitable roles for them. These processes are complex, nuanced, and require the consideration of many factors. The following chapters will review considerations required in the development and evolution of contemporary personnel selection systems with a particular focus on online testing systems.

The current chapter provides a brief overview of common issues encountered by personnel assessment systems in the military domain and acts as a frame of reference for the reader. Chapter 7 illuminates the environment that is driving change in current military selection processes and explores, in-depth, the challenges faced. Chapter 8 provides a detailed overview of the current state of the personnel assessment strategies used by each of the nations participating in the RTG, while Chapter 9 explores risks and mitigation strategies related to the creation of an online testing system, including human factors, and test implementation considerations. Chapter 10 then compares the benefits and costs of developing an advanced online testing system with the use of more traditional systems. Chapter 11 provides a detailed review of considerations when one is contemplating the use of mobile platforms as part of the application and applicant assessment process. And finally, Chapter 12 summarises the antecedent discussion and provides a checklist of considerations that can be used when deliberating the introduction of an advanced selection system.

When the RTG group members started to meet to exchange information, it quickly became clear that commonly used terms, such as *online*, were used differently by some organisations. For clarity, Annex A provides definitions for personnel selection-related terms that are used in this document.

6.1.1 Unique Challenges of Military Personnel Selection

It is important that new recruits can fulfil the requirements of their chosen position and, ideally, that their occupation is a good *fit* for them, so they will thrive in their role and become a valuable member of the military organisation [4], [26], [28]. In contrast with the selection of personnel for civilian occupations, determining the best occupation for an applicant to a military organisation is a uniquely challenging task. Applicants tend to be young and they often have relatively minimal work histories or training that can be used as indicators of their future performance. In addition, military occupations often require unique training or skills that are not readily available to civilians. Consequently, external measures or indicators of applicants' potential performance are difficult to ascertain prior to hiring. Rather, more general measures of ability (e.g., general cognitive ability) or behaviour (e.g., personality measures) are commonly used as predictors.

In addition, the lifestyle of military members is substantially different from that of civilians. Military members are often required to be exceptionally flexible with regard to their working and living conditions, show

considerable resilience in the face of adverse conditions, and demonstrate great loyalty, and physical and moral courage. Consequently, assessments of person-organisation fit in a military organisation are also challenging, as they do not conform to traditional assessments. Thus, military personnel selection systems are faced with a particularly challenging task when trying to identify which military occupation would be best for both the applicant and the organisation.

Failure to determine a suitable occupation for a new military member can lead to dissatisfied and unmotivated soldiers. This can result in attrition, which can, in turn, compromise the strength of the military. In addition, a military member unable to fulfil the demands of their position may compromise the efforts of others, damage equipment, or accidentally harm themselves and others. In the case of nations where military service is mandatory, it is important for the selection process to identify occupations where the conscript can find a meaningful role which they will complete in a motivated manner. In military forces where service is voluntary, good person-job fit is equally important for the maintenance of both individual and team morale, and force capability. Consequently, it is important that military personnel selection systems identify optimal occupations for each recruit, as failure to do so can result in long-term costs for the individual, the organisation, and society in general.

Military selection processes must also work within the laws, policies, and priorities set forth by their government. Diversity and equity priorities must be respected as much as force strength mandates and other political priorities. These priorities often change quickly, thus the selection system must be sufficiently agile to accommodate such changes. For example, just as governments might decide to increase or decrease the size of the military, and thus the rate of enrolment/enlistment because of changing geopolitical or funding priorities, so too might the government decree that the composition of the force (e.g., in terms of gender and racial diversity) must change to better mirror the makeup of the society from which the military is drawn. While some of these requirements might impact private sector companies, others are specific to the public sector, which includes the military. The more important of these current drivers of change are explored in Chapter 7.

6.2 ONLINE PERSONNEL SELECTION TESTING

While the use of computer-based test administration approaches has become commonplace, the use of traditional testing environments has largely endured. That is, test takers are asked to attend a testing centre on a specific date and specific time, in order to complete a computer-based test in a proctored environment. The emergence of the internet, however, has provided expanded options for the administration of personnel selection assessments. *Online* assessments can now be completed at any location serviced by the internet. Applicants can now complete tests in the convenience of their home or other locations at a time that is suitable to them. They may be able to schedule testing such that it does not require time off work or travel to other locations to complete selection tests. As a result, traditional testing resources can be freed up, and the convenience of being able to complete the test at home may result in a larger and perhaps more diverse group of applicants attempting the assessments and ultimately entering the military.

What is gained in convenience in test administration must, however, be weighed against the risks that can ensue from the increased loss of control in the test administration process. The consequences of the departure from a well-controlled and proctored environment must be carefully examined. As a test may be accessed anywhere one can access the internet, it is clearly far more difficult to monitor the test takers. Concerns emerge over issues such as authentication (“is the test taker who they claim to be?”); cheating (“is the test taker using unsanctioned resources to gain advantage, such as internet searches, knowledgeable friends, knowledge of test content, etc.”); and test security (“is the test taker recording test items to develop unsanctioned and uncontrolled copies of the

test and its answer key?”). If the test takers are not monitored when they complete the test online, they are considered to be completing an Unproctored Internet Test (UIT). To help maintain test security and fairness, there are a variety of options to monitor test takers, ranging from active web cameras to the statistical analysis of response strategies. The approaches used by the nations participating in the RTG to mitigate these issues are presented in Chapter 8, while their effectiveness is reviewed in Chapter 9.

6.3 ASSESSMENT IN MILITARY PERSONNEL SELECTION

Each of the military organisations of the nations participating in this RTG have developed programs that present a variety of approaches to the challenges associated with the assessment of people seeking to join the military (Chapter 8). However, as technology advances and societal expectations change, most militaries continue to face the common challenge of how to develop and maintain legally defensible selection and classification programs that are able to process large volumes of applicants in the most efficient, effective, and standardised manner.

6.3.1 Common Standardised Personnel Assessments

Standardised personnel tests or assessments are some of the most effective tools that can be incorporated into the personnel selection process [23], [24]. Such tools have been employed in military selection since the first World War when, in 1917, Robert Yerkes, Leis Terman and David Wechsler developed two selection tests, the United States Army Alpha and Beta tests, to help predict the likely training success of the draftees [6]. Since then, a wide range of assessment tools have been developed to help optimise both civilian and military selection processes, including cognitive ability tests, personality assessments, Situational Judgement Tests (SJT), and the structured collection of biographical data (biodata). As these assessments are referred to in subsequent chapters, their design and implementation will be briefly reviewed here.

6.3.1.1 Cognitive Ability Assessments

Cognitive ability, which is also referred to as general mental ability, has been shown to be one of the more valuable predictors of occupational success in both military and civilian occupations [23], [24]. All nations participating in the RTG incorporate some variation of a test of general cognitive ability in their selection process.

Cognitive ability is a multi-faceted psychological construct, components of which may be assessed through a range of ability tests including, but not limited to, verbal, numerical, and abstract reasoning. These ability tests may require *crystallised intelligence* which requires pre-existing knowledge to solve the test questions, (e.g., how to complete specific math operations), *fluid intelligence* which requires abstract reasoning (e.g., solving logic problems), or both. For a detailed review of cognitive ability and cognitive ability assessment in personnel selection, see [19], [29].

Cognitive ability test items are often formatted as multiple-choice questions with one correct answer. Unfortunately, this item format makes such tests particularly vulnerable to cheating. Test questions and response options may be surreptitiously recorded by test takers thus comprising the security of the test items. The compromised tests items can be reviewed, and the correct answers determined, with these correct answers memorised or recorded and used clandestinely in later testing sessions by others to cheat and improve their test scores. When tests are administered in a proctored testing environment, a critical function of the test proctor is to ensure that no test items are recorded and that no test takers are secretly using a concealed list of the correct answers. It is generally accepted that proctored testing can be employed to promote test fairness and enhance the validity of the test results.

Personnel selection testing is transitioning towards online (internet-based) testing environments, and consequently the ability to effectively proctor testing sessions by traditional means is no longer feasible. When exploring the possibility of online test administration, the ability to fairly administer traditional multiple choice cognitive ability tests in an online environment is a common concern. While the concern is valid in the context of using a multiple-choice cognitive ability test in a high-stakes screening environment (e.g., as a personnel selection tool), there are a number of other approaches to the use of these tests in an online environment that can help attenuate the concern. Common options include using the unproctored test results as a preselection measure with proctored follow-up (allowing the test administrator to compare performance on unproctored and proctored versions of the same test) or providing the unproctored test results to applicants as an indication of how well they may be expected to perform on the proctored test, in order to manage the test taker's expectations.

6.3.1.2 Personality Assessments

Personality assessments typically ask test takers to report the extent to which a series of behaviours, preferences, or opinions are representative of themselves. Often, these are related to the classic *Big Five* personality traits: agreeableness, conscientiousness extroversion, neuroticism and openness to new experiences [20] which have been linked to relevant military performance outcomes [15], [21].

Measures of personality traits are less vulnerable to overt cheating in unproctored testing scenarios, as the test takers are usually unaware of the personality constructs that are being assessed. Test takers may, however, provide socially desirable responses based upon what they think are the desired personality characteristics associated with military service. The problem is that desirable personality traits may be job specific, in that in some situations, scoring high on a particular personality facet (e.g., extroversion) may be beneficial for one occupation, but less desirable for another. Such patterns are not always what a naïve test taker might expect, meaning that an applicant seeking to make themselves more attractive to the selection psychologist might in fact achieve the opposite. Studies have shown, however, that response patterns to personality tests do not differ significantly between proctored and UIT administrations [1], [2], [8], [9]. It is also unlikely that a personality test will be utilised as the sole decision-making tool (i.e., a person is rarely excluded from consideration for not meeting a specified cut-off on a personality construct), meaning that the stakes are arguably less high than might be the case for cognitive ability testing. Therefore, personality assessments generally lend themselves to administration in online unproctored testing environments better than some other techniques.

6.3.1.3 Situational Judgement Tests

SJTs assess test takers' opinions, biases, judgement, or interpersonal skills by presenting challenging social scenarios and asking the test taker how they would (or should) respond (e.g., dealing with a disruptive co-worker [14]). Typically, test takers are presented with a scenario, and they must indicate which of a series of response options is most appropriate or rate the effectiveness of a response that is presented. As the scenarios and the response options are nuanced, there is often more than one reasonable response, and each response may be scored differently [10]. Consequently, without access to the scoring key, it is exceptionally difficult to determine an optimal response strategy for those contemplating cheating. Applicants may of course adopt socially desirable response patterns, as they might with personality assessments, but the unproctored testing paradigm does not provide an inherent advantage here either way. Thus, SJTs also lend themselves favourably to online unproctored testing environments.

6.3.1.4 Biographical Data

Most selection processes require biodata to be collected from applicants. This information reflects an applicant's past behaviour and experiences in a variety of circumstances and situations, and may include data on their

education, work history, skills and specific abilities, and personal and domestic circumstances. The review of biodata to predict job performance has a long history (see [18] for a review) and there is some disagreement in the scope of information that should be considered biodata [17], [18], [25].

Regardless of what approach (theoretical or empirical) is used to interpret the biodata, the collection of such information lends itself to an online approach. There are no opportunities to cheat other than by providing inaccurate information, which can also be done in a proctored environment. If biodata are collected using an online unproctored approach, where applicants may provide the information while at their own homes, there is actually a possibility that the data collected may be superior to what might be collected using other approaches. If candidates provide the information at their convenience, with no acute time restrictions, they may be better able to consider all possible background information that they could provide and potentially remember specific experiences, such as the completion of additional training programs or courses, that they may have not recalled if they were providing the information through an interview or at a recruiting centre. Also, applicants may be able to access specific factual information about a program (e.g., start and end dates of a training program) which might not be readily available if other data collection approaches are used. The provision of the information online might also reduce errors in the classification of various biodata elements. For example, the provision of more accurate information regarding such factors as whether the training utilised formal or informal/self-paced instruction methods, or what level the program was aimed at (e.g., introductory or advanced) allows for more precise classification of the information provided and potentially facilitates automatic scoring, which can result in time and money savings for the organisation conducting the selection process.

Overall, online biodata collection can offer a number of advantages over other biodata collection approaches. When it is used in a standardised personnel selection process, it can help establish each applicant's eligibility, and in some cases their suitability, for military occupations. Biodata can provide a relatively efficient and, if properly implemented, unbiased measure of various psychosocial constructs, and as such they can contribute to the efficiency and fidelity of applicant selection processes.

6.3.1.5 Diversity in Assessments

Traditional cognitive ability, personality, and SJTs, as well as biodata collection, can all provide useful information as part of the military personnel selection process. Some assessments (e.g., personality, SJT, and biodata collection) lend themselves well to unproctored online administration, while others, such as traditional cognitive ability tests, are more vulnerable to cheating and breaches of test security when administered as part of a high-stakes selection assessment using online methods. The remaining chapters discuss a variety of considerations that should be taken into account when developing advanced military personnel selections programs that predominantly involve online testing approaches. It is important to weigh these considerations in relation to the type of selection tool being considered, and the potential costs, risks, and benefits associated with each tool. Overall, however, it is clear that developments in information technology (IT) are now enabling advanced assessment approaches that may provide more effective and efficient assessments than traditional approaches.

6.4 COMPUTER-BASED TESTING

Prior to the emergence of contemporary IT solutions, personnel assessments were commonly implemented in a paper-based format. However, with advances in computer technology, many assessments have been adapted to be displayed on a computer. This allows test items to be presented in a more visually appealing fashion incorporating colour, graphics, animations, and videos. In addition to improvements in the aesthetics of the

presentation, the computer-based administration of personnel selection tests also allowed for practical advances in test security such as presenting test items in a random order for each test taker and eliminating the need for managing numerous copies and versions of test booklets that can be lost or stolen. Using computers to administer tests also allows for automated scoring, which reduces the potential impact of human error and accelerates the selection process. These advantages alone provide a strong case for the use of computer-based tests. However, these approaches alone are simply the digital embodiment of traditional paper-based tests. More recently, advanced tests that take advantage of the power and flexibility of rapidly improving IT capabilities have been developed for use in a selection context.

The introduction of computer-based test administration provides the opportunity to implement more complex and advanced testing paradigms. For example, computers can automatically generate a unique test for each test taker from pools of items with pre-determined characteristics such as item difficulty or determine which item should be administered next depending upon the test taker's responses to preceding items.

6.4.1 Linear on the Fly Test Development

A Linear On the Fly Test (LOFT) is a test created by a computer system that assembles each test at the time of testing. Test items are selected from large pools based on programmed or pre-determined test specifications. Each version of the test generated will have similar psychometric characteristics but will be comprised of a unique set of items [27]. This reduces the exposure of individual test items, which adds to the longevity of the overall item pool; it also improves test security by making it difficult for test takers to cheat by memorising the correct answers to each item, as the item pools usually contain hundreds of items.

6.4.2 Computer-Adaptive Tests

The Computer-Adaptive Test (CAT) paradigm takes advantage of the flexibility of computer technology by creating a test item presentation protocol by which the administration of each test item is contingent upon the test taker's response to one or more previously administered items. In the case of a CAT of cognitive ability, if the previous testing item was answered correctly, the computer will select a *more* challenging item from the test item pool; alternatively, if a test item is answered incorrectly, the test taker may receive a *less* challenging item, or the assessment may be terminated [11], [27]. Through this process, the test taker receives items that are more closely matched with their actual level of ability and are not required to answer items that are far too easy or difficult. Thus, pertinent information on the test taker's ability is determined in a more time efficient manner. The administration of such adaptive tests can only be practically achieved with the use of computers.

6.4.3 Multi-Stage Adaptive Testing

Multi-stage adaptive tests (MSATs) are a variation of a CAT implementation. Test items are presented in small batches commonly called *testlets*, with each testlet having its own psychometric properties. Decisions on which testlet to present next is based upon the test taker's performance on the previous testlet. This contrasts with CAT testing, where item presentation decisions are predicated upon the test taker's performance on the previous item [13], [16].

6.4.4 Test Item Development

All three approaches (LOFT, CAT, MSAT) provide greater test security by increasing the variability in item presentation to test takers and, in some cases, reducing the number of items that need to be presented in order to adequately assess each test taker's ability level. It is important to be aware that the development of tests using

these approaches requires the use of Item Response Theory to establish the item parameters [12]. Establishing item parameters requires the collection of large numbers of observations (i.e., lots of people attempting the item) on the performance of each item at the item pilot testing stage in order to achieve stable measurements. Typically, these approaches also require large item pools in order to be implemented effectively and securely. The extensive pilot testing required by these approaches means that any efficiencies in test administration that might subsequently emerge must be weighed against the increased resource requirements associated with developing the larger item pools.

6.5 CONSIDERATIONS FOR THE USE OF ONLINE PERSONNEL SELECTION TOOLS

There are a number of additional issues that must be taken into account when considering the incorporation of online personnel selection tools. These are briefly outlined below, and further explored in subsequent chapters.

6.5.1 Integration of Online Selection Tests in a Military Recruiting and Personnel Selection System – The Importance of the Recruitment Website

The incorporation of an online testing system has the potential to substantially change the recruiting and selection process. In order for an online selection test to be successfully integrated into a selection program, the online test must be a good fit with the existing military recruiting and selection environment. That is, the test must become a natural component of the selection process. For example, requiring applicants to complete most of the application in-person, or via paper-and-pencil, and then complete one test online might simply add confusion to the process. A well-developed military recruiting website will help serve as a catalyst to make sure the test takers are informed of what is expected of them, and motivated to provide their information and complete the various selection tests to the best of their abilities online.

Today, the need to devote resources to the development of highly functional websites with a professional appearance to support recruiting and selection is well accepted. The use of the internet for recruitment has grown very rapidly in recent years [5], and has profoundly changed the ways in which applicants seek employment opportunities, and organisations manage their recruitment activities [3], [7], [22]. Job posting sites and individual corporate sites aggressively use their web presence to create new and different types of relationships between organisations and candidates [7]. Many militaries are now in direct competition with private sector firms for capable applicants. As many of the in-demand applicants tend to be younger, with a great deal of exposure to technology and cutting-edge websites, poorly developed application and testing software with limited functionality could dissuade prospective applicants from continuing with the application process. Consequently, when considering the implementation of an online testing system, ancillary costs related to maintaining a professional and engaging recruiting site may also need to be taken into consideration.

6.6 SUMMARY

Military personnel selection systems typically process large numbers of applicants each year. As some military occupations frequently require skills or training that are not common in the civilian population (i.e., driving an armoured vehicle, or flying or servicing a fighter jet), applicants are not likely to have the relevant work experience which could reliably be used as an indicator of future performance. As such, military personnel selection units are required to develop or procure alternative selection measures that can help predict future job performance. Fortunately, IT is enabling rapid changes in the quality and manner in which personnel selection

assessments can be administered, as well as facilitating a far greater diversity in the nature of the assessments that can be used. In particular, the omnipresent communication power of the internet allows practitioners to move the domain of assessments from traditional supervised settings to unproctored online administration.

The subsequent chapters review the status of the military selection systems currently used by the nations participating in the RTG and the challenges they face. Options to advance these personnel assessment and selection systems are explored, with both a near-term perspective and a more holistic view of how best to incorporate these advanced emerging technologies to improve the selection process.

6.7 REFERENCES

- [1] Arthur, W., Jr., Glaze, R.M., Villado, A.J., and Taylor, J.E. (2008). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18(1), 1-16.
- [2] Beaty, J.C., Nye, C.D., Borneman, M.J., Kantrowitz, T.M., Drasgow, F., and Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal of Selection and Assessment*, 19(1), 1-10.
- [3] Caers, R., and Castelyns, V. (2011). LinkedIn and Facebook in Belgium: The influences and biases of social network sites in recruitment and selection procedures. *Social Science Computer Review*, 29(4), 437-448.
- [4] Caldwell, D.F., and O'Reilly III, C.A. (1990). Measuring person-job fit with a profile-comparison process. *Journal of Applied Psychology*, 75(6), 648-657.
- [5] Cappelli, P. (2001). Making the most of on-line recruiting. *Harvard Business Review*, 79(3), 139-148.
- [6] Cardona, R.A., and Ritchie, E.C. (2007). US military enlisted accession mental health screening: History and current practice. *Military Medicine*, 172(1), 31-35.
- [7] Charrière, V., Dejoux, C., and Dupuich, F. (2014). L'impact des réseaux sociaux et des compétences émotionnelles dans la recherche d'emploi: étude exploratoire. *Management Avenir*, 2(68), 137-163.
- [8] Dages, K.D., Zimmer, S., and Jones, J.W. (2017). Pre-employment risk screening: Comparability of integrity assessment technology platforms. *International Journal of Selection and Assessment*, 25(4), 390-400.
- [9] Kantrowitz, T.M., and Dainis, A.M. (2014). How secure are unproctored pre-employment tests? Analysis of inconsistent test scores. *Journal of Business and Psychology*, 29(4), 605-616.
- [10] Lievens, F., and Motowidlo, S. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, 9(1), 3-22.
- [11] Lilley, M., Barker, T., and Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education*, 43(1-2), 109-123.

- [12] Lord, F.M. (1980). Applications of item response theory to practical testing problems. Routledge.
- [13] Luecht, R., Brumfield, T., and Breithaupt, K. (2006). A testlet assembly design for adaptive multistage tests. *Applied Measurement in Education*, 19(3), 189-202.
- [14] McDaniel M.A., and Nguyen N.T. (2001). Situational judgment tests: A review of practice and constructs assessed. *International Journal of Selection and Assessment*, 9(1-2), 103-113.
- [15] McHenry, J.J., Hough, L.M., Toquam, J.L., Hanson, M.A., and Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43(2), 335-354.
- [16] Mead, A.D. (2006). An introduction to multistage testing. *Applied Measurement in Education*, 19(3), 185-187.
- [17] Mount, M.K., Witt, L.A., and Barrick, M.R. (2000). Incremental validity of empirically keyed biodata scales over GMA and the five factor personality constructs. *Personnel Psychology*, 53(2), 299-323.
- [18] Mumford, M.D., Barrett, J.D., and Hester, K.S. (2012). 16 background data: Use of experiential knowledge in personnel selection. In N. Schmitt (Ed). *The Oxford Handbook of Personnel Assessment and Selection*, 353-382. Oxford University Press Inc.
- [19] Outtz, J.L. (2002). The role of cognitive ability tests in employment selection. *Human Performance*, 15(1-2), 161-171.
- [20] Rothmann, S. and Coetzer, E.P. (2003). The big five personality dimensions and job performance. *SA Journal of Industrial Psychology*, 29(1), 68-74.
- [21] Salgado, J.F. (1998). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, 11(2-3), 271-288.
- [22] Sameen, S. and Cornelius, S. (2015). Social networking sites and hiring: How social media profiles influence hiring decisions. *Journal of Business Studies Quarterly*, 7(1), 27-35.
- [23] Schmidt, F.L., and Hunter, J.E. (1981). Employment testing: Old theories and new research findings. *American psychologist*, 36(10), 1128-1137.
- [24] Schmidt, F.L., and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.
- [25] Schmitt, N., and Golubovich, J. (2013). Biographical information. In K.F. Geisinger, B.A. Bracken, J.F. Carlson, J.-I.C. Hansen, N.R. Kuncel, S.P. Reise, and M.C. Rodriguez (Eds.), *APA Handbook of testing and assessment in psychology*, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology, 437-455. American Psychological Association.
- [26] Sørli, H.O., Hetland, J., Dysvik, A., Fosse, T.H. and Martinsen, Ø.L. (2020). Person-organization fit in a military selection context. *Military Psychology*, 32(3), 237-246.

- [27] Weiss, D.J. (2013). Item banking, test development, and test delivery. In APA handbook of testing and assessment in psychology, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology, 185-200. American Psychological Association.
- [28] Westerman, J.W. and Cyr, L.A. (2004). An integrative analysis of person–organization fit theories. *International Journal of Selection and Assessment*, 12(3), 252-261.
- [29] Wilhelm, O. (2005). Measuring reasoning ability. In O. Wilhelm and R.W. Engle (Eds.) *Handbook of understanding and measuring intelligence*, 373-392. Sage Publications Inc.

Chapter 7 – MILITARY PERSONNEL SELECTION AND TECHNOLOGY: EVOLVING NATURE OF MILITARY SELECTION

Stefan Hofer

Federal Office of Bundeswehr Personnel Management
GERMANY

Catherine Smith

British Army Recruiting and Initial Training
Command (ARITC)
UNITED KINGDOM

Jan O. Heimdal

Norwegian Armed Forces
NORWAY

Sybille Dunker

Federal Office of Bundeswehr Personnel Management
GERMANY

7.1 INTRODUCTION

Military selection, like any other process in society, cannot be viewed as a static thing. There are many aspects which have an impact on the way military selection is conducted, and selection will need to develop to meet these pressures. It is not possible to describe all the potential influences on selection processes, but this chapter will outline the three main sources of these influences, how they affect military selection, and what considerations will need to be made in future selection processes.

Firstly, military selection processes are influenced from within the armed forces themselves. Section 7.2 describes changes in mission requirements and direction, and how these could affect selection processes. It also highlights the NATO Framework for Future Alliance Operations (FFAO) and demonstrates the impact that this framework has on the Knowledge, Skills, Abilities, and Other characteristics (KSAOs) required to be successful in military selection.

Secondly, military selection processes are influenced by the demands resulting from societal changes. One characteristic of modern society is increased consideration for the experience of minority groups during selection, with the associated aim of increasing diversity and inclusion in the military forces that defend society. As demonstrated in Section 7.3, the need for selection to be fair for minorities has a significant effect on the methods and methodology used in military selection.

The third major influence on military selection comes from the very population that the military is targeting to sustain itself into the future. Section 7.4 shows that applicants put stronger and more varied demands on the organisation they ultimately choose as a potential employer. This changing attitude towards work and working conditions affects many different aspects in the military, particularly the military selection process.

However, it is important to note that these influences on the military selection process not only put demands upon it but can also lead to opportunities for development. Additionally, it is important not to regard these influences in isolation as they are interdependent. Section 7.5. discusses how demands should not be viewed as potential threats, but as possibilities to make selection processes better, to increase fairness, and improve candidate experience.

The military's involvement in military missions is openly discussed in our societies. The feedback that is received, both directly and indirectly, leads to a NATO FFAO that pays heed to public opinion. Influence in the

other direction can be found, too. Circumstances encountered and situations experienced by military members on missions are also openly discussed and can also change public opinion directly or indirectly. There is therefore a reciprocal link between external and internal influences. In addition, it is clear that the attitude of applicants cannot be viewed as independent from changes in society. Changes in opinion can lead to societal changes, and these changes influence the opinions of potential applicants in yet another reciprocal relationship. With consideration of this mutual interplay, the remaining sections take a detailed look at the three major influences.

7.2 INFLUENCE FROM INSIDE THE ARMED FORCES

This section outlines some of the requirements relating to the personnel aspects of future NATO operations, which have been developed from analyses performed in the context of the NATO FFAO. Two different FFAO reports are taken into consideration, one from 2015 [4] and the other from 2018 [5]. In addition, the analysis of these frameworks, outlined in STO-TR-HFM-286, a report produced by NATO HFM RTG 286 (*Leader Development for NATO Multinational Military Operations*), is also considered.

The NATO FFAO reports aim to identify what will be needed for future NATO missions in general, including an analysis of the demands that potential future military contexts and military missions may place on military leaders and personnel. This in turn provides an indication of the future contexts that military selection procedures, including online testing, will have to consider and operate in. These can be described by looking at the KSAOs necessary to enable military personnel to meet the requirements of future operations. They represent a view into the future which is complex and undetermined in many aspects, and which may not transpire with 100% certainty; they tend to concentrate on KSAOs needed for future leaders, not military personnel in general. Nevertheless, it is possible to extrapolate the description of future requirements for leaders to include the wider scope of military personnel, to allow us to form an overall impression of the impact of mission requirements on military selection processes.

NATO HFM RTG 286 presented the results of a gap analysis regarding leader capabilities deemed necessary for future NATO operations, based upon NATO FFAO (i.e., Refs. [4], [5]). Both FFAO reports provide foresights into how Alliance forces might transform to be successful in the future security environment. The gap is described as a framework of leader capabilities required to meet future challenges. These capabilities can be considered as complementary to the existing leadership capabilities typically targeted in today's selection and training.

In the remaining part of this section, a short description of the gap analysis contained in STO-TR-HFM-286 shows how leadership capabilities will need to adapt to ensure military success in future operations. Based upon the analysis of the FFAO 2015, three main characteristics of the challenges put on the Alliance forces can be formulated:

- A need to expand the number and type of partnerships through continuous partner interaction;
- A need for increased operational agility and shared resilience; and
- A need to be more flexible, adaptable, and responsive.

As part of the analysis, leader development is described explicitly in the FFAO 2015 ([4], p. 28):

Leader Development: Future operations will increasingly require military leaders with greater political, technological, cultural and sociological awareness in order to better identify and mitigate risk while capitalising on opportunities. Future military forces will need to foster a culture of technological awareness where its people seek to understand technology and its impact on operations. Military

leaders should be educated and trained in new technologies and capabilities, including autonomous and robotic machines, big data, cyber, and space systems. They should be taught to integrate technology into operations and contribute to the development of new concepts, doctrine and legal frameworks. Military personnel should continue to improve cultural understanding and language skills and train to develop specific regional expertise when necessary. Units should maintain adequate English language proficiency to ensure an ability to communicate clearly across the joint force.

One key aspect highlighted in this summary is the need to increase technological awareness in future leaders, for them to understand the impact technology will have on operations, and for them to integrate technology into their roles. These capabilities support the higher aim described by the challenges put on Alliance forces as described above. They help leaders to achieve greater *operational agility and resilience*, they foster *building new relationships*, and lead in the long run to more *flexible, adaptable, and responsive* forces. The focus on technology here demonstrates how prominent a role it will play in future operations and that it will be crucial for success. For soldiers, greater technological training will be needed, which will require a level of comfort and competence in using a range of technologies.

A more detailed look by NATO HFM RTG 286 (see STO-TR-HFM-286) leads to a more differentiated view on the description of leader development given by the NATO FFAO 2015 [4]. The group identified the following leader capabilities:

- Greater political, technological, cultural, and sociological awareness and language skills.
- Need to understand non-military environments including financial, cultural, ethnic, and religious networks.
- Innovative and creative leaders.
- Establishing trust in new partnership.
- Establishing enduring relationships that will allow commanders to issue mission-command style orders that convey intent.

Again, the need for technological awareness is highlighted. It is very likely that these leaders will need to communicate through technology in order to respond quickly and efficiently to requirements and will need to communicate their intent clearly using online methods. This further demonstrates the strong role technology will play in future operations. As noted above, NATO HFM RTG 286 performed a gap analysis, based upon the proposition that the future will involve a more complex security environment. There will also be rapid and unforeseen changes to this environment which must be anticipated and prepared for, and there will need to be a synthesizing of domains that interact in various ways. This synthesis will be both a challenge and an opportunity to be capitalised upon.

Combined with existing leadership capabilities in the areas of ethics, relationships, and technology, the additional requirements identified by the group (including the need to be ‘credible,’ ‘federated,’ ‘aware,’ ‘resilient,’ and ‘agile’) suggest a number of over-arching leadership capabilities that will be required for the successful prosecution of future operations. These include building capable forces to meet single-nation and alliance challenges; expanding the number and types of relationships amongst allied nations; dealing with increasingly sophisticated and complex levels of information management; building trust; and understanding shared risk amongst partners.

Each of these over-arching or meta-capabilities can be broken down to reflect lower-level or subordinate capabilities. For example, ‘agility’ can be described as comprising a combination several leadership qualities,

including creativity, understanding of context and situation, a mission-command type leadership philosophy allowing for decentralised, flexible decision-making, and the ability to train and operate effectively with multinational units.

Based upon the capabilities identified in the FFAOs, a framework was developed listing the core competencies needed by leaders in order for them to be able to act effectively in the future environment. Many of these competencies can be seen as additional to existing competencies typically trained for in current educational and development programs. As such, they represent areas directly associated with the challenges of the future environments.

Additionally, many of them can be linked to a need for greater technological understanding. For example, expanding the number of types of relationships between nations can be achieved quickly and effectively through using a combination of in-person and online communication methods, and dealing with complex information management will require strong technological skills and understanding of technology information. As the nature of both single-nation and Alliance priorities shift, it is becoming increasingly likely that technology will play a central role in the challenges of the future. It is therefore essential that future soldiers and military leaders can demonstrate confidence in using technological systems and all should be familiar with basic technological processes. Using online testing methods in selection helps respond to this requirement by allowing candidates to use the technology they need be comfortable with, and by demonstrating to candidates that the organisation they are looking to join is technology-based, forward-thinking and ready to respond to a range of challenges. Online testing methods can also provide the flexibility needed to respond to changing demands as required competencies change, to provide the skills needed to meet new or evolving priorities.

7.3 SOCIETAL INFLUENCES

Over the last years, recruiting has become increasingly difficult for both civil and military employers as a tangible pressure in the field of personnel selection has emerged. Two reasons for the rise in pressure can be attributed to changes in society, namely business competition (the ‘war for talent’) and the treatment of minorities, which are described in the following sections. To address these factors, it is necessary to change and modernize certain aspects of the selection methods currently utilised; for example, a move to web-based testing procedures can have positive effects on applicants’ reactions and attitudes towards selection methods and the organisation employing those methods. In addition to the two pre-existing societal pressures noted, the COVID-19 pandemic has placed, and will continue to place, particular requirements on selection processes, which are considered at the end of this section. This section will also demonstrate that the use of online technology in the selection process can help to respond to these influences and be responsive to societal changes and can allow military employers to continue to meet the competencies required as outlined above.

7.3.1 The ‘War for Talent’

While a positively perceived selection method itself plays an important role during the selection process, another important aspect must not be forgotten: the aspect of *business competition*. Military organisations are competing against civil organisations for quality personnel in an increasingly digitized world. In most cases, signing an employment contract with a civil organisation does not usually mean being placed in situations where one might be exposed to physical and mental harm, even death, but being a soldier does carry this risk. Although roles in civilian organisations could be similar to a soldier’s job in terms of the core professional tasks required, the civil employee won’t be required to carry a weapon or to take part in deployments to operational zones. Thus, making the decision to work as a soldier is always linked to aspects and considerations that are not part of other jobs.

Therefore, it is important that the recruitment process places as few barriers as possible in the path of the applicant, and that it is comparable to the recruitment processes used by other organisations that use modern hiring methods. For those who are interested in joining the military rather than a civil organisation, a smooth recruitment procedure and a good candidate experience during selection and testing will ensure that fewer applicants are lost during the recruitment phase through self-selection out of the process. It is the only way that recruitment targets will continue to be met. Online testing methods are becoming widely used in civilian recruitment and military employers will need to utilise similar methods to have a comparable candidate experience. Using these methods can demonstrate to candidates that the employer is forward-thinking, up to date with current methodologies and happy to invest in new initiatives. Online testing can also improve the candidate experience by reducing practical barriers such as the need to travel to a testing location and can offer quicker and more responsive feedback. Using technology to facilitate the recruitment process in this way reduces the number of trigger points which may cause a candidate to remove themselves from the process and can speed up certain elements to ensure that the military offer remains competitive and good candidates are not lost to civilian organisations through long recruitment timelines.

7.3.2 Increasing Diversity

Another important goal for many nations is to broaden the level of diversity within their military organisations. One way to do this is to introduce or enhance the use of quotas in order to increase participation levels of different minority groups. Personnel selection methods tend to show different validity scores for members of certain population sub-groups based upon racial or ethnic, socioeconomic and gender-based factors. This phenomenon is described as the 'diversity-validity' dilemma.

Several ways of reducing this adverse impact have been discussed by Ployhart and Holtz [8]. One way to reduce adverse impact is to use a broader range of predictors, rather than focussing solely on, for example, cognitive measures. This approach enables an applicant from a recognised diversity target group to demonstrate their claims for selection in a variety of ways, which can help overcome the bias that occurs with a single predictor that might be more prone to adverse impact. Another way of reducing this effect is via targeted recruitment, where a potential candidate's attention is drawn to a position for which recruiting strategies have been designed specifically to address adverse impact to sub-groups (e.g., Ref. [6]).

Online testing is another means by which reducing subgroup discrimination might be addressed, by offering a means of encouraging members of minorities to apply for a military career by enabling them to undergo selection in familiar surroundings, and in a blind fashion where ethnicity, disability, gender, or age cannot be initially seen during the testing phase. In some organisations, cognitive testing is conducted in the earlier stages of recruitment and can be used as a sifting criterion. By conducting these tests online in a fully blind fashion, this could ensure that quality candidates proceed to the next stage of recruitment confident that demographic factors have not played a part. This could therefore help to reduce both prejudices on the part of those doing the assessment, and anxiety (due to the risk of being treated unfairly or prejudiced against) on the part of those being assessed, applying for a career in the military. Using online testing can also ensure fair treatment for other aspects of diversity, such as socioeconomic by reducing costs needed to attend testing at a physical location and for disabled applicants who might have practical barriers which would prevent them attending a physical testing location. As society changes and true diversity is expected in all organisations, it will be crucial for employers, both military and civilian, to be able to demonstrate fair and unbiased recruitment processes. The use of online testing can significantly contribute to a positive perception of a fair and transparent process.

7.3.4 The Impact of COVID-19

The transformation process currently being encouraged by pressure coming from business competition and the need to develop a military more representative of the society the military is here to protect, as described above, has been accelerated since early 2020 by the COVID-19 pandemic. In a way unprecedented in the modern world, many countries in the world advised or even ordered their citizens to reduce travelling and limit direct social contact, in order to help reduce the risk of being infected and spreading the virus further. This has increased the pressure for responsible employers, and especially for the military as representative of the governing authorities, to contribute to efforts to contain the pandemic.

One way that military organisations might contribute to this effort is to eliminate, or at least reduce, the need for potential applicants to travel to recruitment/selection centres by increasing the use of online testing in selection. The development of online testing processes and procedures, including the infrastructure needed to implement such procedures, incurs a significant cost for employers, including military organisations. Once these procedures and associated infrastructure have been established, it would not make sense to discontinue their use, even if COVID-19 were to disappear entirely, given the amount of investment that has already taken place, and will continue to occur over the next few years. The move to online assessment and selection that has occurred in response to COVID-19 has highlighted the practical benefits (such as reduced staffing and administration) of using these methods, although the impact of this shift in terms of predictive validity and candidate experience is yet to be fully determined. As a result, it can be seen that the influence of COVID-19 will have a permanent and ongoing effect on military selection processes. The move to online testing during COVID-19 has demonstrated how recruitment can remain flexible and response to unforeseen, global changes. Keeping these systems in place future-proofs military recruitment and ensures that employers remain prepared for the next big change whilst continuing to recruit high quality candidates through any periods of change or uncertainty.

7.4 CHANGES IN THE TARGET POPULATIONS

Driskell and Olmstead [2] listed several fields in which psychology is linked to the military context, one of which is personnel selection. Traditionally, psychological tests were used to assess an individual's capabilities for certain jobs. During the intervening 30 years since that publication, many things have changed with respect to both technological advances available for use in selection and the changing characteristics of the target population. While computerized testing might have been possible and used in only a few organisations in 1989, it has now become a common method for selecting or filtering potential candidates. As technology itself has improved and become more widely available, the attitudes and expectations of the target population have also changed as people become more familiar with technology usage in selection. For most people in NATO countries, electronic devices like smartphones, tablets or notebooks are used daily in their personal lives, and so they are increasingly familiar and comfortable with using the technology for an increasing range of activities.

In addition, most people are accustomed to using these devices to access the internet. People are used to getting any information they desire instantly and in any location, due to the mobile nature of the internet. As they get more and more used to doing a wide range of business and activities via the internet, it is only reasonable that they will expect to do the same concerning their career opportunities. This means not only gathering information about potential career opportunities and job requirements, but also communicating with a possible employer, lodging their application (including all the necessary supporting documentation such as copies of qualifications, biographical data and the like), and otherwise pushing forward their application process as far as possible without wasting time going through the postal system, waiting for bureaucratic letters and requests for other information, and attending a recruitment office.

This increasing use of technology together with the requirement to adapt to the COVID-19 situation described previously, has led to a reasonable assumption that applicants nowadays desire to use their devices not only for submitting their applications, but also for the conduct of the selection process itself (such as conducting an internet-based test via a mobile device). Salgado and Moscoso [9] have found that test takers prefer internet-based testing over paper-and-pencil testing for several reasons, including perceived comfortability and respect to intimacy. A critical review of applicant perspectives during selection demonstrates applicants' positive reactions to web-based selections methods [3]. Since it is getting more and more difficult to address and attract potential military candidates, it is obvious that the expanding use of technology in our societies is suggesting a need to adapt our selection methods to fit these needs, in order to increase the applicant pool and attract new joiners by meeting their expectations in terms of candidate experience.

7.5 RESULTING OPPORTUNITIES FOR DEVELOPMENT

Tippins and colleagues framed guidelines and discussed opportunities and risks related to Unproctored Internet Testing (UIT, [11], [12]). One of the possible opportunities that can result from implementing UIT in a selection process is ability to reach a larger audience. In traditional personnel selection scenarios, a potential candidate is usually required to undertake multiple steps before filling in a single diagnostic test or questionnaire, including making an effort to get informed about their potential job before making the decision to apply only after that do they get the opportunity to be invited to complete a test. The amount of pre-assessment effort currently required of applicants, including a willingness to research the job in question and the conditions of service they will be subject to, the need to travel to and from a recruitment centre, etc., requires a high level of involvement, high costs, and high motivation to progress the application.

By using UIT, this upfront commitment can be substantially reduced, which might help ensure more applicants remain through the recruitment process and reduce drop-outs due to practical issues such as travel. An applicant searching for a job might skip through several web pages by organisations before being 'caught' by an employer whose application process is most informative and leads to quick results. In contrast with a traditional selection process, technology makes it possible to do away with the requirement for an applicant to complete a written application, send it to the employer, and undertake cognitive tests at different points, by making it possible do all of this in one step with UIT.

Braddy, Meade and Kroustalis [1] highlight that applicants react more positively to appealing websites, and Ott and Ulfert [7] conclude that UIT is a simple means of personnel marketing. In the current 'war for talent,' these potential benefits cannot be ignored. Other potential benefits of using UIT, from the applicant's perspective, include:

- Quicker information seeking, making the organisation appear more attractive.
- Reduced waiting time due to the ability to combine several steps in the recruitment and selection process.
- Reduced travel to and from recruitment and selection centres.
- Quicker results arising from the selection process.

In addition to the perceived improvement in candidate experience, UIT can also result in several benefits to employers, including reduced resourcing costs and more upfront information for sorting and filtering candidates.

Other potential benefits include:

- Reaching a larger audience via the internet.
- Reduced administration of personnel selection due to better preselection.
- Projecting an improved ‘high-tech’ contemporary image of the military as a potential employer.
- Being more competitive as an employer.

7.5.1 Other Benefits and Opportunities for the Employer

The use of UIT within the military recruitment and selection process could also result in several other potentially major benefits for those organisations prepared to make a significant investment in and commitment to the technology. One advantage of UIT is its ability to produce a large amount of data (‘big data’) in a relatively short period of time. This can contribute to improving the validity of the selection process in two main ways: the first deals with the process of job and work analysis (e.g., Ref. [10]); the second is more concerned with the ability to adjust selection methods and standards easily and rapidly.

Job and work analysis can be simplified by using computer- assisted test systems. As increasing amounts of test scores and other relevant ‘predictor’ data are collected automatically and stored in a database, more and more information becomes available for researchers to look for correlations between ‘predictor’ data and appropriate ‘criterion’ data. This method of job and work analysis has been described as an ‘empirical and statistical approach based on personal characteristics’ [10]. It potentially offers a relatively simple method of gaining insights into a job, and the characteristics of people doing that job, by identifying correlations between predictors and job criteria. UIT as an automatized process might be an effective first step in tracking and comparing both applicant characteristics and subsequent relevant performance measures, leading to more accurate person specifications and employment profiles, thereby increasing the validity of a selection process.

Such automatized processes might also help improve an organisation’s ability to react quickly when changing or revising test items or tests. Building a database containing both test scores and relevant outcome variables allows the psychometric performance of tests to be continually monitored, which would alert the organisation more quickly to any changes that might be required. While this outcome can also be achieved via a standard computer-delivered testing program, UIT offers the potential to implement this sort of continuous improvement regime more rapidly, leading to changes to tests that can be made in a shorter time frame. This would contribute to maintaining test validity over time and allow for better maintenance of tests used in the selection process.

7.6 CONCLUSION

In summary, this chapter has outlined the three main sources of influence which will direct future military selection processes: influences from within the armed forces due to changes in the nature and direction of future missions; influences arising as a result of societal changes, including the ‘war for talent,’ an increasing need for improved diversity and inclusion, and the impact of the COVID-19 pandemic; and influences from a changing target population, who are increasingly accepting of, and indeed expecting, increased use of technology in selection. Although these influences are already presenting challenges that future military selection methods will need to address, they also provide opportunities for improvements to be made in contemporary recruitment and selection systems and methods. Unproctored internet testing allows flexibility to help face these pressures by providing a testing platform which can potentially be fairer and less biased towards minorities; it can help make the military more attractive to candidates, by representing it as a future-thinking, high-tech employer of choice,

and removing practical access issues to make the process easier and smoother; and it can help improve the flexibility of the organisation to make changes in order to adapt to new requirements as they emerge.

7.7 REFERENCES

- [1] Braddy, P.W., Meade, A.W., and Kroustalis, C.M. (2005). Organizational website usability and attractiveness effects on viewer impressions. In 20th annual conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA, 29, 1-10.
- [2] Driskell, J.E., and Olmstead, B. (1989). Psychology and the military: Research applications and trends. *American Psychologist*, 44(1), 43-54. doi.org/10.1037/0003-066X.44.1.43.
- [3] McCarthy, J.M., Bauer, T.N., Truxillo, D.M., Anderson, N.R., Costa, A.C., and Ahmed, S.M. (2017). Applicant perspectives during selection: A review addressing “so what?, what’s new?,” and “where to next?” *Journal of Management*, 43(6), 1693-1725. doi.org/10.1177/0149206316681846.
- [4] NATO Framework for Future Alliance Operations (FFAO) (2015) [Online]. <https://www.act.nato.int/images/stories/media/doclibrary/ffao-2015.pdf>
- [5] NATO Framework for Future Alliance Operations (FFAO) (2018) [Online]. https://www.act.nato.int/images/stories/media/doclibrary/180514_ffao18-txt.pdf
- [6] Newman, D.A., and Lyon, J.S. (2009). Recruitment efforts to reduce adverse impact: Targeted recruiting for personality, cognitive ability, and diversity. *Journal of Applied Psychology*, 94(2), 298-317.
- [7] Ott, M., Ulfert, A.S., and Kersting, M. (2017). “Online-Assessments” und “Self-Assessments” in der Eignungsdiagnostik. In D.E. Krauss (Ed.). *Personalauswahl*, 215-242. Springer Gabler, Wiesbaden. Doi.10.1007/978-3-658-14567-5_10.
- [8] Ployhart, R.E., and Holtz, B.C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61(1), 153-172.
- [9] Salgado, J.F., and Moscoso, S. (2003). Internet-based personality testing: Equivalence of measures and assesses’ perceptions and reactions. *International Journal of Selection and Assessment*, 11(2-3), 194-205.
- [10] Schuler, H., and Kanning, U.P. (Eds.). (2014). *Lehrbuch der Personalpsychologie*. Hogrefe Verlag.
- [11] Tippins, N.T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2(1), 2-10.
- [12] Tippins, N.T., Beaty, J., Drasgow, F., Gibson, W.M., Pearlman, K., Segall, D.O., and Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59(1), 189-225.



Chapter 8 – ONLINE TESTING EXPERIENCES AND BENCHMARKING SURVEY

Jan O. Heimdal
Norwegian Armed Forces
NORWAY

Jimmi Andreassen
Danish Armed Forces
DENMARK

8.1 INTRODUCTION

There can be a range of approaches to the implementation of an online testing system as a component of a military personnel selection strategy. In order to garner a perspective of the scope of approaches currently used by the nations participating in the task group, a survey was developed to ascertain their current practices and attitudes towards online testing.

This chapter summarises the findings from a survey distributed among nations participating in this RTG in 2018, regarding online tests used for selection purposes. A second survey was distributed in 2021 to check for any updates in test use and development. The surveys were distributed via email, and responses collected online. The questions included in the first (and main) survey are listed in Annex B. A visual depiction of the content captured in the survey is presented in Figure 8-1.

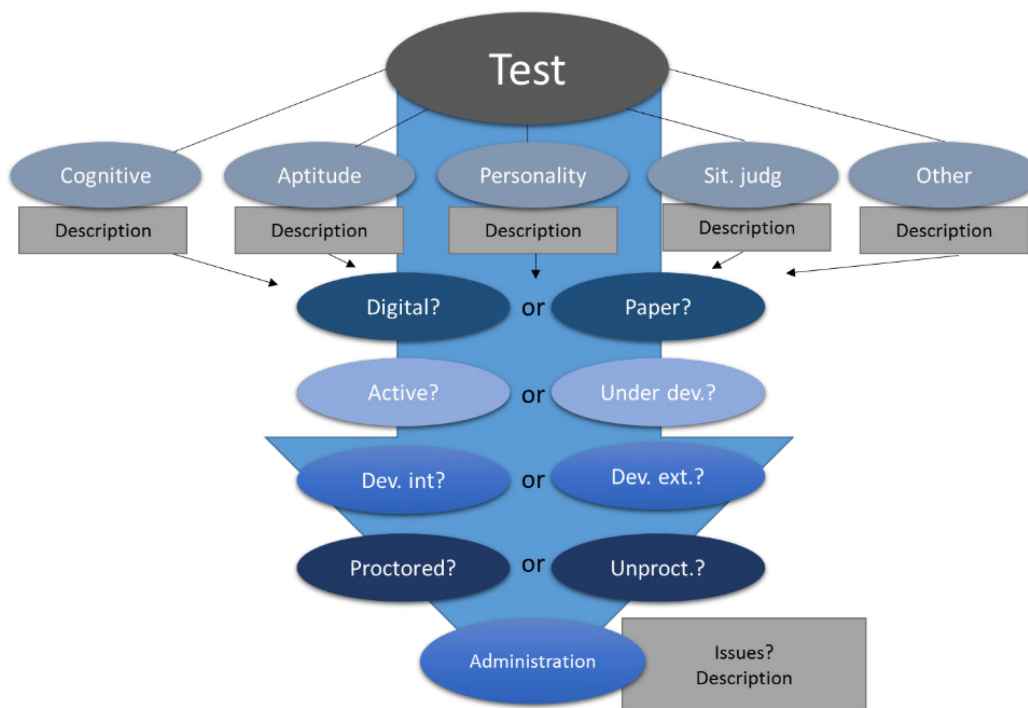


Figure 8-1: Overview of Topics for the Survey. Note: Sit. judg = situational judgement test; Under dev.? = under development; Dev. int.? = developed internally; Dev. ext.? = developed externally; Unproct.? = unproctored.

As depicted in Figure 8-1, the survey asked respondents to report on the following:

- The types of tests used in selection processes of their respective armed forces. Predefined categories (Cognitive, Aptitude, Personality, Situational Judgement or Other) were provided;
- A short description of test;
- Digital or paper format of test administration;
- Whether a test was active (at the time of the survey) or under development;
- Whether a test was developed internally or by an external contractor or vendor;
- Whether test was proctored or unproctored, with a brief description to elaborate; and
- Any issues regarding test administration (e.g., cheating).

The main results of the survey are summarised below. These results are presented with a focus on providing information regarding the development and introduction of online testing rather than on the frequency and purpose of use. This is partly because there was no consistency in the level of reporting across nations. Therefore, the focus is on the practical aspects and experiences concerning the introduction and use of online testing. This information is presented below, grouped by the nature of the tests used (e.g., cognitive, personality).

8.2 COGNITIVE TESTS

Cognitive ability tests most commonly assess applicants' reasoning ability and may require the application of pre-existing knowledge (crystallised intelligence) or abstract reasoning (fluid intelligence), or both, to solve problems presented in the tests [3].

8.2.1 Findings

In the reported survey, representatives of all the 11 participating nations described that they used cognitive ability tests. Further, most nations reported using or having access to more than one cognitive ability test. Based on the information gathered, it appears that what representatives considered or labelled as a test are, in fact, a set of sub-tests measuring different sub-constructs.

The different sub-tests or the various constructs (i.e., skills or abilities) that were listed were:

- Abstract-logical thinking.
- Advanced mathematics.
- Arithmetical problems.
- Attention.
- Basic mathematics.
- Calculation capacity.
- Codes, syllogisms.
- Error detection.
- Figures.

- Literacy test.
- Memory.
- Non-verbal intelligence.
- Non-verbal series.
- Number series.
- Numerical fluency.
- Organisation.
- Perceptual patterns.
- Problem solving.
- Reading comprehension.
- Reasoning.
- Spatial orientation.
- Verbal and non-verbal analogies.
- Verbal classification.
- Verbal logic.
- Verbal reasoning.
- Word similarity.
- Working memory.

The typical general cognitive test consisted of multiple sub-tests that together made up the overall test. The majority reported that their general cognitive ability test consisted of three sub-tests, while others reported having up to 10 sub-tests. Most of these covered the following three areas:

- Verbal;
- Number/mathematics; and
- Abstract/figure/non-verbal.

8.2.1.1 Test Development

Representatives reported that a majority of their cognitive tests had been developed in-house, and a smaller proportion indicated that their tests were developed by a vendor, whereas four tests were reported as being “under development”. Some nations have also developed their own test norms representing their respective populations of interest. In Australia, an online general cognitive ability test developed by external vendors/providers is being trialled; other tests were developed either in-house or by allied military forces. Alternatively, Canada developed a spatial ability test in collaboration with a university, and a working memory battery using in-house resources; both tests are still under development.

8.2.1.2 Test Platform

All cognitive tests reported in the 2018 survey were administered using a Personal Computer (PC) / laptop. None of the tests were available for completion through mobile phones, and only one was available for completion

through a tablet. Some nations reported administering their cognitive tests on paper as well, but to a smaller portion of the total numbers of applicants. Only one nation used both, the online and unproctored testing option. But applicants would then also have to complete a proctored verifying test. The majority of nations distribute their tests digitally, not online, in a proctored setting.

8.2.1.3 Test Administration

The nations described their testing administration process, which included information on how candidates are invited to take tests, how the test taker's identity is authenticated, at what stage in the recruiting process the candidate completes the test, and whether or not feedback is given to the candidate on their performance. As expected, the responses to this varied; below are descriptions provided by different nations, which give some insight into the different ways test administration is conducted.

Norway enrolment and classification, three General Mental Ability (GMA) sub-tests, proctored, digital:

The GMA battery is a proctored test, starting with an instruction from a selection officer. The test is taken in step two of conscript assessment, where approximately one third of the cohort meets before a draft board. Feedback is given in a recruitment and placement interview later the same day, as the test score is a selection criterion in itself. Those who do not get feedback or are evaluated unfit for duty (based on the GMA or other selection criteria, e.g., medical), can, at any time, contact the main HR division and get feedback.

Czech Republic, battery of different tests, proctored, digital:

Candidates subscribe using an online form, after which they are asked to come to a recruitment office, where the applicants are given a short briefing, their background is checked, and a date for testing at the military hospital is assigned. At the military hospital, they undergo approximately 2.5 hours of testing and an interview with a psychologist. The interview discusses their motivation, anamnesis, and their test results. If there is only a small failure, additional tests are administered. If a person has been tested before, different tests are administered, and more focus is given to previous experience. After finalising all examinations, they are invited to complete the physical tests. The final step is an interview with an army personalist [HR personnel] who will assign them to a position within a unit based on the testing assessments.

Sweden, Cognitive ability battery of 10 tests, digital, proctored:

After screening by an online self-report questionnaire and basic military recruitment screening (for formal requirements), candidates are called for basic military admission testing and examinations, which take place over a whole day. On this day, a proctored cognitive test battery is administered. Feedback is then given by a psychologist during a counselling session.

8.2.1.4 Test Security and Cheating

On the topics of cheating and test security, a number of different issues, practices and insights were reported. Some main points are summarised here:

Germany: online, proctored.

As stated by Germany, they do not have any concerns over candidates cheating and test content security, because the test is presented in a proctored environment, and over a secure network and there is no possibility to copy results.

Belgium: digital, proctored.

Testing instructions describe what is allowed and not. Candidates are observed, and their identity checked. They have individual workstations. If cheating, the person can be excluded from the selection for a period of 8 months.

Norway: digital, proctored.

Norway has mandatory service and has experienced ‘faking bad’ in order to avoid service. However, as the candidates are tested after having completed an online survey in advance, and then invited to the testing based upon their replies and scores on this prescreening, the concern is minimal.

United Kingdom (UK) – British Army: online unproctored, followed by online proctored.

Applicants complete an online, unproctored pretest to qualify for a proctored, digital test at the assessment centre. They report no concerns about security as the test is item generated, and verified, so there is little benefit to the candidate to cheating (under unproctored conditions).

Denmark: online unproctored, followed by onsite proctored.

Denmark has almost the same practice as the UK with an unproctored GMA pretest that qualifies applicants for the next step in the assessment centre, which entails a verification test, or another GMA test administered in a proctored setting.

New Zealand: paper, proctored.

On the topic of future approaches towards security and online, unproctored testing, New Zealand offered the following proposals. Future development might consider the voluntary use of the candidate’s webcam to verify them or a phone call to them as they begin testing (to reduce risks). Technological solutions in future are likely to incorporate reaction time analysis, fakeability warnings and adaptive testing systems – all of which have some researched links to the reduction of fakeability. In addition, regular validation may identify if faking is damaging predictive validity of testing.

Australia: online, unproctored (under development).

A description of the administration process and test instructions regarding cheating for a test under development was provided by Australia. Clear warnings are delivered up-front, at the commencement of the online testing session, stating that a) anyone called cheating will be excluded from consideration, and b) proctored verification testing designed to detect cheating would be conducted for applicants when they attend the Recruiting Centre. Verification testing will be implemented for a proportion of applicants (approx. 30%) reaching the ‘Assessment Day’ – the final margin for error between the initial online test result and the result of the verification testing (i.e., the marker for determining whether the two scores are significantly different) has initially been set at 2 x SEM (Standard Error of Measurement).

Sweden: digital, proctored.

Sweden offers insight and reflections regarding learning effects. As the test is not adaptive, there learning effect of test items affecting re-testing. Problems with learning effects are minimised by an imposed time limit for possible re-testing (no re-testing within 18 months, or similar). A related problem is that the testing material might be known to the candidates as it could have possibly been copied earlier and practised.

Canada: digital, proctored.

Applicants to the Canadian Armed Forces first complete an online application to establish their basic eligibility (e.g., minimum age, citizenship, education, and security questions). The information is reviewed and if the requirements are satisfied, the applicant is asked to come into a recruiting centre to complete a proctored cognitive ability and as well as a personality test that are administered in a digital format. The Canadian Armed Forces is in the process of creating an online unproctored cognitive ability test to be used as a pre-selection measure.

8.2.1.5 Planned Improvements in Test Administration

Some nations indicated and described both dissatisfaction and satisfaction with their test administration, along with an indication of what they possibly wanted to change. In general, the number of nations that indicated they were satisfied was about same as the number that were not satisfied or wanted to improve their testing administration.

Australia described the need to develop new online tests to make testing more accessible to applicants, thus increasing the likelihood of attracting applicants who might otherwise have reservations about attending a Recruiting Centre. This was based on the awareness that some prospective applicants (e.g., minority groups) applied to join the military in smaller numbers than potential applicants from majority groups.

New Zealand described the need to move from paper-based testing to digital testing, and listed possible benefits of this transition which included:

- Widening of applicability and availability of selection methods to a far greater range of trades;
- Validation of all selection methods (predictive validity);
- Reduction in the costs of selection;
- Improved weighting methodology/battery composition;
- Introduction of new tests;
- Use of methods to reduce faking and adverse impact; and
- Smarter hurdling (cutting using the best predictor of performance first).

Sweden described the need for more flexible testing solutions and a general need for updates. They indicated that long testing procedures might be tiring for the candidate, which might affect performance. In addition, this long proctored testing restricts the usability of the test since it is not advisable for “quick” screening purposes, or for testing online. They proposed to collaborate with others on a new fully adaptive cognitive test battery.

8.2.2 Summary: Cognitive Tests

The use of cognitive tests made up the main portion of reported testing among the nations. Although most nations have their own developed tests and different structuring of tests (how many sub-tests there is in a test), most nations use GMA tests with (at least) a verbal, numerical, and abstract component; others have other possible subtest categories in addition.

Most nations use digital, not online, and proctored tests. Still, a number of different, possible benefits of online testing were described.

8.3 PERSONALITY TESTS/INVENTORIES

As described in Chapter 6, personality tests are typically used to assess peoples' motivations, preferences, interests, and style of interacting. Personality self-report inventories typically ask applicants to rate their level of agreement with a series of statements designed to measure their standing on relatively stable personality traits.

Personality is described using a combination of traits or dimensions. The personality traits model most widely used is the Big Five or the Five-Factor Model (FFM). The popular NEO-Personality Inventory-Revised (NEO-PI-R) assesses the five dimensions of Extroversion, Emotional Stability, Agreeableness, Conscientiousness, and Openness to Experience. However, other personality measures also exist and are used.

In Schmidt, Oh and Schaffer's meta-study [4], the authors suggest that certain personality characteristics together with GMA account for a significant proportion of job performance. Traditionally, the trait of Conscientiousness is to a larger degree associated with performance.

8.3.1 Findings

Most nations have one or more personality tests available. The most reported type of test was based on the FFM. In addition, a few other tests were reported that more directly targeted specific constructs like "social life" (Norway) or leadership (Sweden). Also, a few other general tests like Cattell's 16PF (UK and Czech Republic) and the *Persönlichkeits Stil und Strörungs Invetar* (PSSI; Personality Styles and Disorder Inventory) (Czech Republic).

8.3.1.1 Five-Factor Model

A number of different FFM tests, which differed in both the number of items and facets included, were mentioned. Some nations used different versions of the FFM tests for different purposes. One example of this is France, which uses a test called the OCEAN, with 77 items and two facets per dimension, as well as another test, the NEO 3, with 240 items and six facet per dimension. The latter is largely used for special assignments, like Special Forces selection. Another example is the Czech Republic which uses both the NEO-PI-R and the NEO Five-Factor Inventory.

Most nations reported that their FFM tests were developed in-house, whereas a few were commercially developed (e.g., by Hogrefe). There were a few examples where tests have been "borrowed", or inspired by other branches, with a focus on the internal validation of such tests. In addition, two nations have tests under development, in-house.

All tests are distributed digitally and proctored. Most nations also have the opportunity for a paper-based testing.

The FFM personality tests are typically used to support an interview, and/or viewed in combination with other tests and data to make an overall judgement about an applicant. There were no reports that clear cutoffs were used for the personality test results, without considering other methods.

Even though all testing takes place in a proctored environment, one central aspect regarding personality tests concerned impression management or faking bad/good on the test. This is one of the reasons test results are used in combination with other methods, and that it can be the subject for discussion during the interview.

Regarding satisfaction with the tests, and possible plans for improvement, a few different opinions were given. Some concerned the need to include measures like values and commitment (France), and the lack of ownership of the data and running costs (when an external provider delivers the test (Sweden)).

8.3.1.2 Other Personality Tests

In addition to the FFM tests, a few other personality type tests were reported. These include the “Commander trait inventory” by Sweden; and the 16PF, the PSSI (Inventory of personality styles and disorders) and N70 (70-items neurotic symptoms checklist) by the Czech Republic; as well as the 16 PF used by UK Army. In Australia, relevant personality constructs are assessed via an interview with a psychologist; they do not use personality tests as part of the initial recruitment/selection battery.

8.3.2 Summary: Personality Tests

Most reporting nations use some version of a five-factor model personality test and primarily distribute this test digitally. The tests are used for different purposes, and one nation might have different versions, but it is typically accompanied by other selection methods, like interviews.

8.4 OTHER TESTS

As described above, the majority of tests reported were cognitive or personality tests. In addition, some different ‘other tests’ were reported which included aptitude tests, Situational Judgement Tests (SJTs), or knowledge tests. The distinctions between these were not very clear, but typically, these tests were used for the selection of aircrew, or individual tests used to measure specific skills or knowledge. For example, SJTs provide complex social scenarios and have test-takers either rate the effectiveness of a response to the scenario or choose the most effective response from a list of alternatives [1] to assess test-takers’ opinions, biases, judgement or interpersonal skills. See also Chapter 6 for a description.

8.4.1 Findings

Seven different aptitude tests, two situational judgement tests, and seven specific knowledge tests were reported.

Belgium reported on an aptitude test under development that measures organisational abilities. A military context is presented with different documents containing descriptions of vehicles, maps, and schedules. The candidate must read this information and then answer 25 questions that refer to this information. The skills that are measured are reading comprehension, the ability to synthesise, to make priorities. They also reported using a situational judgement test that measures leadership abilities. Validation studies will decide how the test will be implemented and used.

Germany reported using four tests that largely measured specific knowledge. These included the electrotechnical knowledge and understanding test, the Räumlicher Vorstellungstest (RVT) measuring spatial ability, the Mathematik-Kompetenzen-Test (MathKom) for mathematical skills, and Mechanik-Kenntnisse-Test (MKT) for testing mechanical knowledge. All these tests have existed since before 2008 and are used as part of a larger selection test battery. Some were developed in-house and others by a vendor. All are digital and still in use.

8.5 UNPROCTORED INTERNET TESTING

A central part of the survey was to evaluate the participating nations' views on the use of unproctored internet testing, particularly their concerns about it, its potential benefits, and nations' experiences with this form of testing.

8.5.1 Concerns About Unproctored Internet Testing

Many indicated that they have concerns over the use of unproctored internet testing. The main concerns are in regard to fraudulent behaviour, authentication of the candidates, the risk of compromising tests, fairness (e.g., applicants having different access to devices and internet, different experience with technology, and different computer skills). All participating nations pointed out that there is a risk that test content can be copied and shared with others or published on the internet. Furthermore, they expressed concern that the applicant can use unauthorised help such as the use of the internet to search for the answers or use the help of a friend or even let someone else take the test. Hacking and general cyber security is an issue that was also of concern. There seemed to be less concern regarding negative impact to the predictive validity of unproctored tests, and costs of implementation and operation of the unproctored testing system.

Access to good internet connection was also mentioned as an important issue. Even though internet connectivity and speed has improved immensely in the recent years, this is not always the case in remote areas in the larger countries. As unproctored internet testing requires good internet access, not ALL applicants will have the same access to unproctored internet testing. To ensure that no applicant is disadvantaged, measures should be taken to ensure that these applicants have the option to complete testing at a testing facility or other site with good internet connectivity. Often unproctored internet testing is used to sort out applicants before inviting them to the testing facilities. Hereby, the opportunity to "recycle" an applicant is made more difficult. For example, when testing is completed in a physical setting, those who fail on tests required for their preferred occupation can be encouraged to try for a different occupation or education.

From the viewpoint of the applicant there are some other concerns as well. Some applicants might have concerns about data security. The lowered opportunity for personal interaction with organisational contacts might be an issue for some applicants, impacting their views of the military organisation. Even though the use of modernised unproctored internet tests will present the organisation as modern, there is also the risk that some candidates will find the use of these tests alienating. Many big private enterprises and organisations, however, make use of unproctored internet tests and, therefore, the main applicant pool, consisting of primarily young people between 18 – 30 years of age, might come to expect these procedures. Research shows that applicants in a selection context have positive perceptions of internet-based testing, and concerns in relation to privacy of information seem more prominent when social media is used as part of the selection process in screening and selecting prospective job applicants [2].

8.5.2 Benefits of Unproctored Internet Testing

One of the main arguments for the use of unproctored internet testing that many of the participating nations mentioned was that it is more efficient and reduces waiting times and travel costs for candidates. The majority also saw it as a way of engaging potential candidates differently so as to reach a larger audience.

Sweden indicated that the use of unproctored tests saves time in the selection process and enhances flexibility. It gives Sweden the opportunity to sort out many candidates in the screening phase, meaning that the quality of those applicants entering the basic military admission testing and examinations are better. This form of screening

might portray the Swedish Armed Forces as a competitive authority, and as an organisation with a competent selection process. Sweden acknowledged, however, that it might be a drawback that questions and answers might circulate among applicants and friends, in discussion groups etc., which might influence impression management over time. This is why they reported that the need for proctored onsite testing is still important as an alternative or as the main selection step after initial screening.

In general, the use of unproctored testing was seen as a possibility to reduce the attendance time of applicants in the selection centre, which in turn would result in a reduction of working hours of personnel required in the selection centre. Therefore, unproctored internet testing can be a significant time and money savings measure for the defence forces. Once unproctored tests are implemented, it was felt that it would be easier to reach a large number of applicants without any extra effort.

8.5.3 Experiences with Unproctored Internet Testing

The UK, Sweden, Norway, Netherlands, and Denmark were the only participating nations at the time of the first survey in 2018 with experiences using unproctored internet testing.

The UK uses an unproctored test on information processing capacity across areas of verbal reasoning, numerical reasoning, spatial awareness, visual speed, and accuracy. It is used as a screening tool and is completed at home if the candidate has access to the internet. For candidates that do not have internet access, they are invited to complete the test at their local Army Careers Centre. Candidates who successfully make it to the assessment centre, will have to go through a longer proctored version of the test, administered under exam conditions. The UK reported no concerns about security as the test is automatically item generated, and verified, so there is little benefit to the candidate to cheating under unproctored conditions.

Denmark has been using a similar procedure since 2016. A GMA test is used as a screening tool. The applicants have to score above a certain cutoff before they are invited to the testing facility for further testing and interview. The cutoff is not fixed but can be changed from year to year depending on how large the applicant pool is, and on how many eligible applicants are needed. When the applicants come to the testing facility they have to go through another GMA test. The two tests used are very different but have high inter-correlation. The first one is a newly developed test from 2017, whereas the second one is an older test used in the Danish Armed Forces since 1957. The results on the second test are weighted higher than the results on the first test.

Denmark conducted a small follow up study after the first year of using the unproctored GMA test. The cutoff the first year was set deliberately low (stanine 3). After finishing the recruitment process, the results from the unproctored test were compared with the results from the proctored test. It showed that if the initial cutoff had been set higher (stanine 4), 80% of those applicants who would have been screened out by using the higher cutoff, would have been false negatives, as these 80% actually scored better on the second test at the testing facility. This result indicates that candidate malpractice might not be as big an issue as many otherwise fear. It seems that the unproctored test setting enhances the risk of candidates underperforming. There could be several reasons for this: the candidates might take the test less seriously and they might be more easily distracted. In the proctored testing facilities, it is possible that the test condition increases the production of stress hormones which enhances individual performance.

Denmark also uses a personality test from an external provider that has been in use since 2014. The personality test is not used for screening candidates but is used to save time in the testing facility and to give applicants the flexibility to choose when to take the test. Validation of test scores occurs through the selection interviews.

In Sweden, all conscripts go through an unproctored screening procedure (to assess formal requirements and to gather personal information) followed by onsite testing and a selection interview. The results from the screening process are used for de-selecting those unsuitable for basic military training, mainly due to health-related issues. Results may also be used for information during interview, and basic military medical/physical examinations. Applicants to special selection (i.e., pilots) are not subject to unproctored testing for the time being but need to have passed conscription service with approved grades. Sweden is currently (2021) considering implementing the use of unproctored testing for general intelligence and personality for screening purposes.

Norway also uses unproctored self-reporting on a number of different variables for the first round of the enrolment of conscripts. The questionnaire goes out to the entire cohort of applicants. The different factors reported on include health issues and a shorter version of the NEO-type of questionnaire. The results are used to qualify candidates for the second round of the enrolment process at selection centres where medical, physical and GMA-testing is carried out. In 2019, Norway implemented the use of an unproctored personality test, a 240 item NEO-type questionnaire distributed through a link to a web page sent by an email to the participants. The results of the personality test are stored in the candidates' electronic military profiles and are evaluated by a psychologist and then used and validated in combination with other sources of test or selection data (e.g., interview) throughout the selection procedure.

Even though Australia, at the time of the survey in 2018, did not use unproctored tests, a review had identified that the number of 'rub points' (i.e., points in the recruitment process at which the prospective applicant was required to physically attend a Defence Force Recruiting Centre [DFRC]) borne by each applicant was quite large. These were viewed as a major disincentive by a proportion of applicants, especially those from targeted diversity/minority groups (e.g., females). In response, commercial service providers / test developers were asked to identify or develop an online approach to testing, under which an applicant might complete testing in the comfort of their own home rather than attend the DFRC.

8.6 DISCUSSION

Even though the majority of the nations represented in the surveys do not have hands-on experience with unproctored internet testing, there is a tendency against using unproctored internet tests on the basis of assumptions that are in line with the findings in two articles by Tippins and colleagues [5], [6]. Tippins [5] explained that many experts and I-O psychologists are wary of unproctored internet testing because of questions about their effects on the standards for acceptable testing practices. One of main points in these papers is that it is never acceptable to use unproctored internet testing in high-stakes situations without some sort of verification test. Another point made is that neither the extent of cheating nor the effects of cheating deterrents in unproctored internet testing is known.

As reported above, participating nations have many concerns regarding test security and fraudulent behaviour and those risks must be taken seriously. However, none of the participating countries that currently use unproctored internet testing reported having any concrete negative experiences in regard to security or fraudulent behaviour. The question remains whether nations should abstain from exploring the opportunities of unproctored internet testing due to risk avoidance, or whether potential gains can be realised through investigating its use on a smaller scale, where the test results are not given the same weight as proctored test results.

Fraudulent behaviour is an inherent risk in almost all selection procedures, and therefore the risk itself should not be the main reason for not exploring the opportunities of unproctored internet testing. The use of unproctored internet tests needs a well-thought-out plan concerning data security, integration of the test with other testing

processes in the organisational structure, authentication of candidates and cheating, protection of test content, as well as strong communication with applicants to ensure that their performance reflects their ability under a setting where many factors cannot be controlled.

8.7 NATO RTG-290 2021 SURVEY UPDATE

The work of the RTG-290 overlapped with the COVID-19 pandemic, a period during which nations had to adjust their selection procedures to adapt to altered rules and guidelines regarding the pandemic. This situation potentially pointed to a higher need for online tests, and/or a higher emphasis on developing these types of tests. As new solutions were identified and explored, it was necessary to capture any new experiences and new issues. For these reasons, another survey was distributed to nation representatives of this RTG (see Annex C). Specifically, representatives were asked to report on how the COVID situation affected their respective nation's selection processes. Three following aspects of their experiences were explored:

- 1) Kinds of challenges that nations had experienced;
- 2) Possible solutions generated to these challenges; and
- 3) Any experiences with using the solutions that were generated.

8.7.1 Findings

Responses were gathered from a total of nine nations: Australia, Belgium, Canada, Czech Republic, Denmark, Germany, The Netherlands, Norway, and Sweden. All of these confirmed that selection had been affected by the COVID situation.

8.7.1.1 Challenges

The challenges seemed to pertain to two different topics: 1) General challenges inflicted by the pandemic; and 2) The effects of these challenges for selection.

The general challenges, overall, had to do with restrictions regarding travel (e.g., lockdowns, restrictions), safe distance, the number of people that could be present at the same time in a particular facility, and disinfection requirements. These limitations made meetings and gatherings pertaining to entire selection processes harder to organise and conduct. These practical restrictions, in turn, resulted in fewer selection assessments being conducted, including fewer administrations of proctored tests even though there was increase in the number of applicants due to the economic consequences of COVID. Even if onsite testing was possible, it was an ethical challenge to make a decision about having applicants travel to the test facility, possibly increasing their exposure to infection.

8.7.1.2 Solutions

In addressing the challenges presented above, a number of different solutions or alterations to existing processes were presented. These solutions were introduced in a number of different areas.

Some solutions were highly practical. These included using larger rooms, or the outdoors, to maintain distance, or having a smaller number of people or groups in parallel processes. Also, reorganising the sequence of activities so that activities requiring a higher level of contact, and thus possibly higher exposure to disease, was at the end of the process, where there is a fewer number of applicants who are still part of the selection process.

One line of solutions involved the use of different kinds of digital and online platforms for information sharing and interviewing, as well as for testing. Some made use of unproctored online tests, and later verified the results in a proctored setting. Some nations used web-cams to conduct proctored testing. Digital and phone interviewing replaced face-to-face interviews. Some supported the decision to conduct interviews digitally with a risk analysis of candidates, focussing on risk factors like mental health symptoms, poor social integration and compliance with the law, academic underachievement, unstable employment history, and other information to evaluate whether or not someone should be interviewed digitally. Others used data like test scores, biographic data, and other information about the candidate to support the interview, whereas others introduced new tests that were digital.

Overall, the solutions to a large extent entailed making use of technological tools and methods, as well as restricting the number of people that were required to meet for selection activities by relying on other or existing data.

8.7.1.3 Experiences with Solutions

The solutions to the challenges presented above were not necessarily ideal, or previously tested and evaluated. Therefore, the nations were asked about their experiences with the different new or altered selection processes. Both positive, negative, and unknown experiences were listed.

Some of the listed experiences concerned the quality of the selection. For example, some indicated that the true effects of changes made to selection during the pandemic were not yet fully available. Such comments reflected the notion that there may be a risk when deleting tests that have been proven predictive. Some indicated that the new situation put higher work demands on selection staff. In addition, some experienced a big turn down (i.e., decline) of invitations by applicants to complete online tests as compared to onsite testing.

On the more positive side, experiences also reflected new insights or benefits. For example, some indicated that the changes resulted in more effective processes, where the different assessment tools played a more defined role. Also, some reported successes in using unproctored ability testing followed up by a verification test, such that only 3 – 5% of applicants needed to complete a full retest due to inconsistent results between the unproctored and verification test results. Further, some reported that webcam proctoring of tests worked at the level of testing of individuals. In addition, a few nations mentioned that an increased acceptance of and even a push towards developing online solutions and selection processes.

In sum, the experiences seem to indicate that exactly how the quality of selection is affected by changes introduced in selection during the pandemic is still not clear. Future validation studies and operational feedback will be needed to provide insight. A lot of the enforced changes were considered to be challenging from a professional decision-making perspective. In other words, the risks of not being able to use tests proven to have predictive validity had to be weighed against opportunities to introduce and use (new) technical solutions.

8.8 CHAPTER SUMMARY

Based on findings from the initial survey of nations (pre-pandemic), all nations use cognitive tests, typically consisting of several sub-tests. Most of these are digital, but only a few online, and some have a paper format available. Most tests are proctored. A number of personality tests are used, most of which are versions of the “five-factor model”. In addition, some other personality tests are reported. In addition to the cognitive and personality tests a few other tests were reported. These fall under the category of aptitude and/or situational judgement tests, and typically assess some specific skill or knowledge. When asked about unproctored tests, only a few nations reported using such tests, along with efforts to counteract the weaknesses of this kind of testing.

A second survey following the pandemic explored the challenges and solutions associated with selection during pandemic. A number of practical and methodological issues for selection were identified. The need for new solutions was also identified, with the use and development of technology being central to this approach.

8.9 REFERENCES

- [1] Lievens, F., and Motowidlo, S. (2016). Situational judgment tests: From measures of situational judgment to measures of general domain knowledge. *Industrial and Organizational Psychology*, 9(1), 3-22.
- [2] McCarthy, J.M., Bauer, T.N., Truxillo, D.M., Anderson, N.R., Costa, A.C., and Ahmed, S.M. (2017). Applicant perspectives during selection: A review addressing “so what?,” “what’s new?,” and “where to next?” *Journal of Management*, 43(6), 1693-1725. doi.org/10.1177/0149206316681846
- [3] McGrew, K. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, 37(1), 1-10.
- [4] Schmidt, F.L., Oh I., and Shaffer, J.A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research finding [Working Paper, Research Gate]. doi.org/10.13140/RG.2.2.18843.26400
- [5] Tippins, N.T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2(1), 2-10.
- [6] Tippins, N.T., Beaty, J., Drasgow, F., Gibson, W.M., Pearlman, K., Segall, D.O., and Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59(1), 189-225.

Chapter 9 – RISKS AND RISK MITIGATION STRATEGIES IN UNPROCTORED INTERNET TESTING

Annemie Defranc
Belgian Defence
BELGIUM

Geoff Gallas
Department of Defence
AUSTRALIA

9.1 INTRODUCTION

Advances in personnel selection testing emerging from technological developments can bring significant benefits to an organisation. Such advances are usually accompanied by a revised risk management framework designed to optimise the benefits and minimise the risks associated with the new types of tests (e.g., test standardisation; proctored delivery). However, when new developments are required by societal or other factors at a much faster rate than expected (e.g., outbreak of war, pandemic), risk management strategies might not always keep up with real world requirements. A perfect example is the COVID-19 pandemic, where many organisations felt the pressure to adapt their selection processes by offering applicants the opportunity of taking web-based personnel selection tests online in their homes rather than requiring them to report to a recruiting centre to be tested.

Introducing high-stakes selection testing in an unsupervised situation introduces a new set of risks, in areas such as test security, applicant identification, cheating, test standardisation and applicant reactions, that are were much less problematic in a proctored test setting. These concerns are identified in the seminal article *Unproctored Internet Testing (UIT) in employment settings* [17]. Such concerns can lead to problems for the organisation conducting the testing, including inaccurate assessment of an applicant who behaves dishonestly, issues of test security, validity and reliability, and associated legal problems. Since the publication, progressively more strategies to deal with these risks have been developed [16]. This chapter focusses on the risks associated with UIT and the strategies that might be employed to help mitigate those risks.

9.2 TYPES OF RISKS

9.2.1 Psychometric Risk

In a supervised selection setting, the test environment is standardised and the identity and behaviour of candidates are controlled. This means all applicants are treated in identical fashion, minimising the risk of later challenges to non-selection by unsuccessful applicants, and optimising test validity and reliability. In an unproctored setting, this standardisation and control disappears, which threatens the reliability and the validity of the test results.

9.2.2 Threats to Reliability

The environment in which a test is administered can have a significant impact on applicant performance on that test. Good testing practices require that applicants take the test in optimal conditions that facilitate performance [17]. An unproctored setting can undermine test reliability due to unstandardised test administration conditions. For example, a candidate who chooses to take a test in an environment full of distractions or in a setting where they can be disturbed during testing, is likely not to do as well as if they would have attempted the test in quiet,

controlled circumstances. In other words, their performance on the test in the unstandardised condition will be underestimated; the test result would not be a reliable estimation of the construct measured on the test and hence not directly comparable to the test results of other applicants who attempted the same test in different conditions [18]. To increase the reliability and comparability of different applicants' test results test instructions should point out that the test should be taken in a quiet environment as free as possible from distraction. However, in the case of UIT, this ultimately remains the responsibility of the candidate.

Whether or not noise and other distractions play a significant role in test performance is inconclusive. However, Tippins and associates reported that only a very limited number of candidates reported being distracted by noise during testing in an unsupervised context [17]; another study even found that candidates reported experiencing less noise in an unsupervised environment than in a supervised environment [18]. While it might therefore be argued that a free choice of test environment is unlikely to significantly impact test reliability, the risk remains that allowing an applicant to attempt a test in an unproctored environment might well open the door to legal and other challenges from those who are unsuccessful.

Aside from the test environment, technical issues such as lack of stability in hardware, software, and the quality of the internet connection can also threaten the reliability and validity of the online test result. Variations in hardware (e.g., screen width) and the speed of the internet can impact performance on cognitive tests involving with images and/or where speed in answering the question is taken into account. Tippins however, noted that many of the technical problems with UIT are not substantial concerns for the majority of applicants [16]. Although stability of internet connection and the consistency of test experience across different equipment (hardware and software) were once considered substantial concerns, many of the technical problems associated with UIT have tended to disappear as more and more people gain access to better technology [16]. Nevertheless, for applicants who don't have access to stable internet, or are not able to take the test under ideal circumstances, organisations should retain the ability to offer the tests to applicants via a selection/recruiting centre. By doing so, one major source of threat to the psychometric quality of the test results (i.e., reduced reliability) is effectively managed.

9.2.3 Threats to Validity

When candidates take a test at home or in another unproctored setting, they clearly have the opportunity to cheat. There are several ways of cheating; examples include using prohibited materials (such as calculators, dictionaries, internet, or stolen test answers), asking others for help, or even asking someone else to take the test in their place. In all of these circumstances, test scores will not reflect applicants' actual ability in the construct being measured [12]. More specifically, there is a risk that the applicant who claims to be completing the test will be less capable than their test score would suggest, potentially resulting in a person who is not fit for the job being hired or enlisted. This can have a follow-on effect, whereby the less capable person ends up taking the place of another applicant who would have been better suited to the demands of the job.

When applicants cheat on a selection test, the validity of the test result is compromised. Usually, three different but inter-related (they all involve some form of deception) types of threat to validity are identified [12], [16], [17]. These involve test security, applicant authentication and cheating, and are discussed in the following sections.

9.2.3.1 Test Security

The first validity threat arises when the applicant has gained knowledge of the correct answers to the test prior to attempting the test itself. This can occur when the test content becomes public, perhaps due to the test being

stolen or ‘leaked’, or where the applicant has completed the same test already on one or more occasions and becomes familiar with the test items due to overexposure [12], [16], [17].

Test security breaches can occur when the test is hacked. In this kind of deception, the test’s content is stolen (either physically or electronically) and shared with others, or published on the internet. This can result in a test needing to be revised or replaced in its entirety, which is a very costly and time-consuming exercise. The storage of test items, scoring keys/algorithms, and respondent information must there be maintained at a very high level of security, protecting the information from unauthorised access and dissemination.

Secure data transfer and information storage processes must be enabled along with the test administration software, preferably via a fire-walled proxy; an independent test database separating test and applicant information from the broader internet must also be maintained. Data must also be encrypted while in transit to guarantee data security. The latter measure will not prevent interception during data transfer, it does however ensure that any data that is intercepted is not interpretable to the person who receives it.

Ensuring the security of test items while they are actually in use is more complex. As the test items are visible on the applicant’s computer monitor during administration, the opportunity exists for the applicant to copy those items via either electronic or manual means. To prevent this, it is possible to disable the Print Screen function, but there is no protection against taking pictures of the screen. It is also possible for a candidate to assume multiple identities in order to keep practicing with the test, until they feel confident enough to take the test ‘for real’ [16]. Possible solutions to this problem (and the following risks) are discussed in the section on risk mitigation strategies.

9.2.3.2 Authentication

The second potential threat to validity arising from the use of UIT relates to authentication the applicant. In other words, how can one guarantee that the person completing the test and providing the test results is actually the person who was invited to participate in the online assessment? It would be a simple matter for the applicant to give their login data to someone who, in their mind, is more capable, and asking that person to do the assessment for them [16], [17]. Without authentication, it is impossible to ensure that the obtained test scores actually correspond to the person who is meant to be doing the test.

9.2.3.3 Cheating

Unproctored internet testing creates the opportunity for the test taker to cheat by using materials that are prohibited by the testing protocol (e.g., internet, dictionaries, calculators, illegal scoring keys) to help them work out the correct answers, or by asking for assistance from others.

According to Wollack and Frammer [19], the extent to which candidates are likely to cheat depends on five factors:

- Stakes – cheating is more likely to occur if the stakes are high (which is often the case in a recruitment and selection context).
- Predisposition – applicants differ in how willing or resistant they are to cheating, and this depends on their personal moral/ethical standards.
- Need – applicants who feel confident that they will succeed in the selection test, will be less likely to cheat.
- Opportunity – do the test conditions and test design allow applicants to cheat more easily?

- Consequences – what happens to someone who is believed to have ‘cheated’? Will they just be forced to take a retest, or will they be excluded from the selection process entirely? Cheating is less likely to happen if the applicant is not convinced that the benefits of cheating outweigh the disadvantages of getting caught.

Wollack and Frammer point out that, of these 5 factors, the testing organisation really only has any influence on opportunity and consequences [19]. Limiting the opportunities for applicants to cheat and highlighting the negative consequences of cheating are the key risk management strategies that will be discussed in the section on risk mitigation strategies.

A final point to note is that these risks: test security, authentication, and cheating, are mainly associated with tests where there is one correct answer to a test item such as cognitive ability or knowledge tests. In the case of personality tests or autobiographical questionnaires there are no correct answers, so deception is not usually a problem. In both cases, in a high-stakes situation such as personnel selection, the applicant is likely to present him/herself in a socially desirable manner, this can occur in both supervised and unsupervised test administration conditions [2], [3], [5], [9], [10]. The first empirical study to measure the validity of biodata, sales potential and conscientiousness (personality) tests in supervised versus non-supervised environments was conducted by Beaty and associates [3]. The results of this study showed that, in terms of validity, applicants responded to these non-cognitive assessments in a similar fashion in both supervised and unsupervised environments. Studies examining the validity of cognitive assessments across proctored and unproctored settings are too limited and varied to draw any firm conclusions about the use of unproctored testing to assess cognitive ability [8].

9.2.4 Reputational Risks

Besides psychometric risks, organisations might be concerned about applicant perceptions. An earlier chapter indicated that the use of new technologies can improve the organisation’s high-tech image, which could increase the attractiveness of the organisation for applicants. However, internet-based procedures can also have a negative impact on some applicants’ perceptions, such as: applicants who might be reluctant to apply to an organisation that makes high-stakes decisions based on tests that are prone to fraud [16]. These negative perceptions can have serious consequences, including legal contest of the equity of the selection process and, potentially, a decrease in the size of the recruitment pool. In addition, if an organisation relies almost completely on internet-based testing for selection decisions, some applicants may form the impression that the employer is not sufficiently involved in the selection process, and is indifferent to their needs. Others might also be concerned about their privacy and data security, with data being transferred over the internet.

Consequently, at a minimum, organisations should ensure that the technology that is used for internet testing is secure in protecting the test content, scoring keys and applicant response data. Besides this basic requirement, it would be prudent either to choose unproctored tests where cheating is less of an issue (e.g., some non-cognitive tests), or in the case of cognitive tests, to implement strategies (discussed below) to minimise the risks.

9.3 RISK MITIGATION STRATEGIES

9.3.1 Non-Technical Risk Prevention Strategies

One of the more basic strategies in the prevention of cheating is to appeal to the responsibility and liability of the applicant. For example, prior to testing, the candidate could be asked to sign a statement confirming his/her identity and committing to honest test-taking. Building in this type of accountability in testing procedures has been shown to discourage some applicants to commit fraud or cheat [11].

Another option is to inform the applicant about the possibility that their scores can/will be verified in a later stage in a proctored test environment (e.g., using an equivalent or shorter test version). Instructions can also point out the negative consequences, should an applicant be caught cheating at any stage. These strategies use of various types of warning instructions has been shown to an effective mechanism to deter applicants from faking or responding in a socially desirable manner in personality testing [6]. Such non-technical strategies aim at encouraging honest test-taking behaviour and/or discouraging candidates to cheat.

Unproctored test results can also provide information to the test taker. Beyond acting as a screening tool which only benefits the hiring organisation, the results from unproctored tests may also be used to provide beneficial information to the applicants of the selection process. If the test scores contribute to establishing eligibility for specific military occupations, the results of unproctored testing may be used to advise applicants of their potential eligibility for a range of occupations, some of which they might not have previously considered. The information on eligibility provided through the unproctored test results may prove to be valuable to applicants and applicants may be motivated to obtain a valid score to inform their own careers choices.

9.3.2 Technical Risk Prevention Strategies

There are several approaches an employer might undertake, from a technical perspective, to discourage or prevent cheating. The first category off such strategies focuses on preserving the security of the test content, by preventing overexposure to and/or copying items and response alternatives:

- **Passwords and logins:** These should be used to protect access to the test and unnecessary exposure of test items [17].
- **Test design:** Drasgow pointed out “Any internet test that administers the same set of items in the same order to all examinees is asking to be compromised. At the very least, the items should be administered in a randomized order. It would be better yet to sample items from a reasonably large item pool.” ([17], pp. 203-204,). CAT (Computer Adaptive Testing) or LOFT (Linear On the Fly Tests) are test development options that can be used to serve this goal. This kind of testing essentially produces a test by retrieving items from a larger pool, with item selection governed by applicant ability [13]. With the CAT and LOFT approaches, there is little value for an applicant to attempt to memorise or otherwise keep track of items, because each respondent will receive another variant of the test.
- **Test response format:** Test security can be enhanced by altering the presentation of the item response alternatives. This involves presenting the applicant with response alternatives in a sequential, random order, such that the applicant has to indicate whether each alternative is correct or not before seeing the next sequential alternative. The response alternatives for an item are presented to the applicant until they have identified the correct response or has made an error. In this way, not all response alternatives have to be presented, which decreases exposure of the response alternatives.
- **Technical precautions:** These include disabling certain keyboard functions such as the copy (Ctrl-C) or Print Screen (PrtScn) functions. However, as mentioned previously, there is no way to prevent someone from physically taking pictures of the items presented on the screen.

The second category of technical strategies focus on preventing different forms of cheating (e.g., using prohibited materials or asking assistance from others) and addressing identity authentication:

- **Browser lockdown:** to prevent an applicant from using prohibited materials to help them determine the correct answer to a test item, the system should be configured such that the applicant cannot start another application, move from one application to another, or access the internet during the test.

- **Test design:** a cognitive or knowledge-based test can be designed in such a way that using prohibited materials such as the internet, dictionaries and calculators provides little practical advantage. When items require insight (such as abstract reasoning items) to be solved, looking up possible solutions on the internet can be difficult and too time-consuming to be beneficial. This is especially the case if there is a time limit per item [1].

These two categories of strategies shown above do not entirely prevent the possibility of cheating. Applicants are still able to ask assistance from others, or ask someone else to take the test for them. One way to prevent this form of cheating is via webcam monitoring. Three approaches can be considered.

One way of using a webcam is through live proctoring, where candidates are continuously monitored by proctor via a webcam feed as they attempt the test. The applicant is first asked to undergo a visual identity check, where their face (as it appears on webcam) is compared with a legal document containing a photo (e.g., a national ID-card, passport, driver's license) that has been supplied previously or is also presented through the webcam. Applicants are then continuously monitored during the entire duration of the test to ensure that they do not cheat, and no one else completes the test. Also, it can be made possible for the proctor to interrupt the test if they notice signs of cheating such as that the candidate has telephone contact or there is unusual background noise [7]. The advantage of this approach is that it makes it possible to control the identity and behaviour of the applicant, and to respond to possible security problems in real time, as would be the case in a traditional proctored setting. The disadvantage lies in that a proctor is needed throughout the entire testing session, which means that test administration depends upon both the number of proctors available (in the case of bulk recruitment, this might be a lot) and the availability of the proctor to observe testing at the required time usually within the organisation's office hours. This can clearly limit the flexibility of UIT in offering testing at any time that is convenient to the applicant. Another disadvantage is that this process may be more resource intensive than in-person proctoring of a group of applicants.

Webcams may also be used to record test sessions for later view proctoring. In this process an identity check is completed as described above, and the applicant's behaviour while completing the test is recorded through the webcam for analysis at a later time. This offers two advantages over live proctoring. The candidate can take the test anytime, and testing is not constrained by the availability of a proctor. It is also more resource-efficient for the organisation in that analyses of recorded videos can be restricted only to those candidates who completed the test successfully. The recorded information can also be viewed at an increased speed. The disadvantage of this approach is that the proctor cannot respond to any detected security issues in real time [15].

In addition, webcams may be used for the purpose of interviewing candidates before and after the online test. During the interviews, the applicants must prove their identities and may be asked about the testing session. This approach is less secure than the other methods mentioned. Objective self-awareness theory [14] in connection with the impression management theory [3] would predict the interviews will reduce cheating as the test takers will focus on themselves and behave in a more socially desirable manner.

One additional consideration when using a webcam is that it could create an extra threshold or hurdle for the applicants to actually participate in the selection procedure in the first place, because of the practical/technical complexities involved, and due to possible concerns regarding privacy.

Microphone monitoring: If a webcam is not available, the microphone on the test taker's computer or cell phone can be used to help monitor the testing session. Test takers could be asked to sit somewhere quiet and keep the microphone on during the online assessment. As speech production on the part of the test taker is not normally required in the testing situation [7], the active microphone can potentially identify sounds not typically associated with proper test-taking behaviour which can be followed up either in real time or later through a review of the recorded session.

9.3.3 Detection Strategies

Two strategies commonly used to detect cheating post-testing, and are focused on identifying test security breaches that do not involve review of test data from a particular applicant are:

- **Web patrols:** This involves scanning the internet on a regular basis to search for sites or forums that are known to provide information on the relevant tests, test items, or responses [4].
- **Dashboard:** This involves monitoring the performance of test items on a regular basis. Sudden increases in correct responses to items can indicate a test security issue, perhaps in response to test items having been posted on the internet or made available to applicants in another way. Compromised items can then be identified and replaced in the item pool.

The following two approaches can be used to assess the likelihood that an applicant has cheated, and require an analysis of an applicant's test data:

- **Data forensics and key-stroke analytics:** This approach which analyses response patterns across test items to detect aberrant responding [4]. For example, if an applicant responds correctly to all items or responds at a much faster rate than what can be expected on the basis of the psychometric properties of the test itself (e.g., an applicant who responds correctly to more difficult items with the same speed as they respond to easier items), this might suggest that the applicant is aware in advance of what the correct responses actually are. Conversely, unexplained delays in the speed with which an applicant responds to simpler items (i.e., extremely long response times) might suggest that the applicant is using prohibited materials (such as an internet search engine on a personal device) to help them respond to test items, or even copying the items for other potential applicants. Similarly, remote monitoring (either in real time or after the fact) of an applicant's key-strokes can indicate whether the applicant is attempting use prohibited functions such as 'Print Screen'.
- **Verification test:** This requires applicants who succeed on the unproctored test to take an equivalent, but often shorter test, in-person in a proctored environment later in the selection process. This serves to check the consistency of scores obtained from unproctored testing. The preferred option is to utilise CAT or LOFT testing approaches for both the unproctored and proctored test administrations, so that applicants are presented with essentially the same test, but do not receive exactly the same questions the second time. If the test score obtained in the supervised setting is significantly lower than the score arising from the unsupervised setting, the applicant's score may be considered to be aberrant, and gives rise to questions on the validity of the score obtained in the unproctored setting [15].

Caution is required with the use of data forensics and verification testing. Conclusions about cheating that are based on these methods are susceptible to errors of Type-I (erroneous detection of test fraud) and Type-II (failure to detect test fraud). As these methods call an applicant's honesty into question and can have serious consequences such as being excluded from consideration for a place in the military, the decision criteria used to indicate cheating must be very accurate, and require a mechanism to exclude other possible causes of aberrant score changes [16]. Even when aberrant responding is detected through these methods, it cannot be automatically interpreted as evidence of cheating [4]. There is a need to investigate the issue with sensitivity, to rule out other possible reasons for aberrance (e.g., physical or psychological factors across both test administrations). However, the completion of a follow-up verification test in a proctored environment is considered a superior approach to increase the validity of test results [15], [16].

9.4 SUMMARY AND CONCLUSION

This chapter provides an overview of the major risks associated with adopting an Unproctored Internet-based Testing (UIT) paradigm/approach for military selection, and identified various strategies that might be used to reduce the likelihood of cheating, and/or detect and minimise the impact that cheating might have on a country’s selection processes.

Table 9-1: Summary of Risks, Risk Prevention and Detection Strategies.

Risk	Prevention	Detection
Authentication problem	Applicant must show ID	Verification testing
Getting unauthorised assistance during testing	Honesty contract; Warning Instructions	Webcam
Use of prohibited materials	Browser lockdown Test design (abstract reason items; maximum response time)	Verification testing Data forensics Webcam
Test security	Passwords and logins CAT, LOFT Item response format Technical precautions	Web patrols Dashboard

As there is limited published research on the effectiveness of online proctoring solutions in addressing inherent security concerns [7], it is impossible to advocate one strategy over another. The primary goal of this chapter was to raise an awareness of the risks associated with UIT and possible strategies to mitigate risk. Readers are encouraged to consult other resources (e.g., Refs. [7], [17]) for fuller coverage on these topics.

9.5 REFERENCES

- [1] Arthur Jr, W., Glaze, R.M., Villado, A.J., and Taylor, J.E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment*, 18(1), 1-16.
- [2] Beaty, J.C., Nye, C.D., Borneman, M.J., Kantrowitz, T.M., Drasgow, F., and Grauer, E. (2011). Proctored versus unproctored internet tests: Are unproctored noncognitive tests as predictive of job performance? *International Journal of Selection and Assessment*, 19(1), 1-10.
- [3] Bolten, H.G., Mummendey, H.D., and Isermann-Gerke, M. (1983). Die Theorie der objektiven Selbstaufmerksamkeit im experimentellen Vergleich mit der Impression-Management-Theorie und dem Bogus-Pipeline-Paradigma. *Bielefelder Arbeiten zur Sozialpsychologie*, 98, Bielefeld: Universität Bielefeld, Fak. für Soziologie.

- [4] Burke, E. (2009). Preserving the integrity of online testing. *Industrial and Organizational Psychology*, 2(1), 35-38.
- [5] Dages, K.D., Zimmer, S., and Jones, J.W. (2017). Pre-employment risk screening: Comparability of integrity assessment technology platforms. *International Journal of Selection and Assessment*, 25(4), 390-400.
- [6] Dwight, S.A., and Donovan, J.J. (2003). Do warnings not to fake reduce faking? *Human Performance*, 16, 1-23.
- [7] Foster, D. (2009). Secure, online, high-stakes testing: science fiction or business reality? *Industrial and Organizational Psychology*, 2(1), 31-34.
- [8] Goodrich, S., and Loh, E. (2021). Unproctored internet testing: Overview of considerations (Director General Military Personnel Research and Analysis Scientific Report DRDC-RDDC-2021-R042). Ottawa, ON: Defence Research and Development Canada.
- [9] Hogan, J., Barrett, P., and Hogan, R. (2007). Personality measurement, faking, and employment selection. *Journal of Applied Psychology*, 92(5), 1270.
- [10] Kantrowitz, T.M., and Dainis, A.M. (2014). How secure are unproctored pre-employment tests? Analysis of inconsistent test scores. *Journal of Business and Psychology*, 29(4), 605-616.
- [11] Lerner, J.S., and Tetlock, P.E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, 125(2), 255-275.
- [12] Lievens, F., and Burke, E. (2011). Dealing with the threats inherent in unproctored Internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology*, 84(4), 817-824.
- [13] Luecht, R.M. (2005). Some useful cost-benefit criteria for evaluating computer-based test delivery models and systems. *Journal of Applied Testing Technology*, 7(2) 1-31.
- [14] Silvia, P.J., and Duval, T.S. (2001). Objective self-awareness theory: Recent progress and enduring problems. *Personality and social psychology review*, 5(3), 230-241.
- [15] Slusky, L. (2020). Cybersecurity of online proctoring systems. *Journal of International Technology and Information Management*, 29(1), 56-83.
- [16] Tippins, N.T. (2009). Internet alternatives to traditional proctored testing: Where are we now? *Industrial and Organizational Psychology*, 2(1), 2-10.
- [17] Tippins, N.T., Beaty, J., Drasgow, F., Gibson, W.M., Pearlman, K., Segall, D.O., and Shepherd, W. (2006). Unproctored internet testing in employment settings. *Personnel Psychology*, 59(1), 189-225.
- [18] Weiner, J.A., and Morrison, J.D. (2009). Unproctored online testing: Environmental conditions and validity. *Industrial and Organizational Psychology*, 2(1), 27-30.

- [19] Wollack, J.A., and Fremer, J.J. (2013). Introduction: The test security threat. In J.A. Wollack, and J.J. Fremer (Eds.), Handbook of Test Security, 1-13. New York, NY: Routledge.

Chapter 10 – BENEFITS AND COSTS OF INTERNET-BASED TESTING

Geoff Gallas and Kelly Battocchio

Department of Defence
AUSTRALIA

The typical scenario is that you go to a large room during business hours and take a written examination along with hundreds of other candidates. A month or so later you receive a letter informing you of your test results and indicating that you may be called at some point in the future for an interview. The whole process seems impersonal and often leaves the prospective candidate wondering exactly where they are in the selection process. From the employer's point of view this type of process is slow and results in a loss of the best candidates who often accept a job offer made by another employer before the governmental agency can even notify them of their test score ...

There were many advantages to using the internet-based test ... applicants are able to take the test at their convenience 24 hours per day, seven days per week, from home or at work. This flexibility negated the need to assemble everyone in one location and employees did not have to take time away from work to take the test ... The results of the test were available to the administrator in real time allowing them to quickly make selection decisions. ([7] p.43)

10.1 INTRODUCTION

Psychological testing is an acknowledged key component of best practice in the field of recruitment and selection [2]. Psychological tests, particularly those measuring general cognitive ability, have consistently shown reliability and validity in predicting job performance (e.g., Ref. [10]). Test results provide a recruiter or an assessing psychologist with an additional, valuable piece of data they can use in making selection decisions. Cognitive testing also provides a cost-effective means of reducing the size of the applicant pool, and hence reducing costs, by screening applicants before they get to the interview stage. This benefit is even more pronounced with the use of internet-based testing, in that there is no time or travel cost for the organisation to arrange for the applicant to attend a testing session in person.

10.2 POTENTIAL BENEFITS OF PSYCHOLOGICAL TESTING

The traditional model of testing, which involves testing sessions being supervised closely by a trained test administrator or 'proctor' (as described in the preface to this chapter) has several well-known benefits. Proctored testing environments remove or control for many of the random sources of measurement error that might afflict or influence the outcomes of psychometric testing (e.g., extraneous noise; too bright or too dim lighting; inconsistent test instructions; inconsistent test timing; uncomfortable seating), thereby allowing meaningful interpretation of the test results [3]. This in turn provides a solid defence against appeals from unsuccessful applicants regarding non-selection, and potential legal action from disgruntled sections of society (or individuals) alleging unfair selection processes (i.e., all applicants are treated the same way).

10.2.1 Benefits of Computer-Based Testing

With computer-based testing, which can also be administered in a proctored environment, there are more benefits still. For example, Computer Adaptive Testing (CAT), described in more detail in Chapter 6, can save both candidates and test administrators time and resources by reducing the number of items to which the candidate is required to respond; candidates spend less of their own time seated before a computer screen. With computer-based testing, test administrators are able to provide concise, accurate feedback to the candidate within a much-reduced timeframe compared to traditional paper-based testing.

10.2.2 Benefits of Internet-Based Testing

Unproctored internet-based testing, on the other hand, while perhaps lacking the control (offered by proctored testing environments) over factors that could impact measurement error also has some advantages, some of which were flagged in the preface to this chapter, including the flexibility this approach offers for test takers, and the speed with which test results can be provided back to the administrator. This flexibility means, for example, that the test can be made available to someone who would not otherwise have time to attend a proctored testing session during work hours. Such an advantage offers the potential for the organisation to be able to increase both the size and the diversity of the applicant pool.

A related advantage is that internet-based testing can improve applicants' view of assessments as a positive experience [1], [6]. The use of internet-based aptitude diagnostics, for example, can result in applicants viewing assessment as a positive experience, for a variety of reasons. Applicants who would not otherwise choose to participate in a traditional selection process might be happy to complete an online assessment instead, which may boost their interest in an employer and increase the probability of them actually applying for a job. Particularly in military organisations, which offer a great variety of occupational specialties and jobs, low-threshold information is of great value to someone who is interested in such a career move but not entirely hooked. Through the use of internet-based tests, within the framework of a self-assessment, an interested person can examine their own motives and interests online, thus learning more about themselves and what kind of jobs might potentially suit them. Any advantages that might be realised, however, will need to be balanced against the risks associated with the introduction of unproctored testing – these risks are described in greater detail in Chapter 9 of this report.

10.3 POTENTIAL COSTS OF PSYCHOLOGICAL TESTING

So, how much does it cost to introduce a new selection testing regime? There are too many variables at play to provide a definitive answer to this question that will be equally accurate in every country across the world. Moreover, the actual monetary costs at each stage will of course vary depending upon what the country in question already has in play. Rather than attempting to answer this question, the following section instead identifies some of the factors that influence the cost of a selection test, whatever the platform.

The section begins with an explanation of cost-related factors that are relevant across different types of testing and testing platforms (i.e., proctored vs. unproctored). It continues with a description of specific costs that exist in proctored environments, and identifies additional costs introduced by the use of computerised testing formats in these environments. Finally, the section finishes with a description of the costs associated specifically with the use of unproctored testing environments. A summary of the most common types of cost encountered is included in Table 10-1.

Table 10-1: Cost-Related Factors for Different Forms of Tests.

	Proctored (Pencil-and-Paper)	Proctored (Computer)	Unproctored (Internet)
Test Development <i>(Note: Costs vary depending on whether using an in-house or contracted solution)</i>	Yes	Yes – including need for larger item pools	Yes – including need for larger item pools
Research	Yes	Yes – including test equivalence research	Yes – including test equivalency research
Staffing Costs	Yes – test administrators (including scoring and report writing)	Yes – test administrators (however, scoring may be computerised)	Yes – IT support
Staff Training	Yes – training administration, scoring and report writing	Yes – training in administration	Yes – training in administration (required for verification testing and/or as a back-up)
Test Environment (Maintenance of Physical Location)	Yes	Yes – must be capable of holding computer hardware/software	Yes (required for verification testing and/or as a back-up)
Security	Limited	Yes – fail-safing	Yes – verification testing
Applicant Time	Testing and travel time	Testing and travel time	Testing time only
Materials	Printing costs (paper forms, etc.)	Computer hardware/software costs	Webpage/Internet

10.3.1 Common Cost-Related Factors

Across both proctored and unproctored testing environments, psychological tests have costs for:

- 1) The test developer or publisher (which may be the employer, in the case of in-house development);
- 2) The employer or test user; and
- 3) The applicant or test taker.

10.3.1.1 Test Development Costs

For the test developer, the most notable costs are associated with test development, monitoring and maintenance, and research (including test norming and validation). A key consideration here is whether an employer wants to develop a selection test in-house or to use a contracted solution. Purchasing test authoring software and developing ‘in-house’ a suite of psychometric tests might be seen as an attractive option by larger employers in particular, in that this kind of software is generally less expensive than purchasing fully-developed tests, and allows the greatest flexibility in developing and changing test items and the like over time [5], [7]. As Mooney points out, however, if you choose to develop a test yourself, you become fully responsible for ensuring that the test is valid and reliable, and that it retains those qualities over time. Having another company undertake this

latter task (test validation) for you is an option, but one that is often difficult (due to access to data, security concerns, etc.) and time-consuming (especially in a military organisation) and *will* cost money.

Maintaining the required capability (test development expertise, psychometric skills, programming ability, etc.) in-house can also add significantly to staff- and employment-related costs, and might be a luxury that smaller militaries cannot afford, especially as such specialist skills are not always available within the local employment market, and might come at a high price [3], [4]. Hinton also points out that trying to maintain an online testing system in-house might also require you to maintain direct responsibility for the development and maintenance of anti-cheating and other user-verification tools and techniques, which once again can add significantly to the costs associated with online testing [3].

If, on the other hand, the preferred option is to contract out all or even part of this process, Mooney points out that different pricing structures, corporate practices, and governance requirements, even laws, can make price comparisons between different companies difficult [7]. One test developer/publisher might charge highly for developing a test, but then offer that test back to the customer at a relatively low ‘per test’ fee, which might be favourable for a large organisation recruiting large numbers of people. Another might employ a different model, such as charging no up-front development costs at all, and seeking to recoup costs via higher test administration fees and a long-term contract with the client (i.e., such as a semi-exclusive ‘partnering’ arrangement). Many test publishers, however, are likely to offer customers some form or level of discount for purchasing their product in bulk.

10.3.1.2 Employer and Applicant/Test Taker Costs

For the employer, there are a number of additional costs. These include costs associated with the development of policies and procedures as part of a wider testing and assessment framework, and costs in training staff in interpretation of results and report writing (and, in cases where tests are administered in a proctored environment, training in test administration).

For the applicant, there are very few costs that are consistent across all kinds of testing, as the different testing platforms and environments offer very different applicant experiences. One consistent cost, however, is the ‘time cost’, or the time taken to undertake the test. The actual time cost depends heavily on three factors:

- 1) The testing environment (i.e., whether the cost is simply the time taken to complete the test, or includes time to travel to a location for the test);
- 2) The technology (e.g., are test results available to recruiters and decision makers immediately); and
- 3) Organisational policies and procedures (e.g., requiring applicants to return to the recruiting centre to complete additional assessment components at a later time).

10.3.2 Discretionary Cost-Related Factors

10.3.2.1 Traditional Proctored Pencil-and-Paper Testing

While this chapter is focussed primarily on exploring the additional costs typically encountered by the introduction of computer-based testing and unproctored test environments, it is not the case that a traditional, proctored, pencil-and-paper test environment has only the baseline costs and no more. For example, with any proctored testing environment, there are costs associated with establishing and maintaining a physical test location of a suitable size and with appropriate physical security measures in place. Additional costs are associated with the purchasing or production of test materials, including answer sheets and profile forms

(for pencil-and-paper testing), administration and technical manuals and any other relevant materials (including, for instance, pens and pencils). Additional staff training is required for staff to administer the tests, which itself adds to the cost, and there are also ongoing costs related to employing the test administrators themselves (e.g., salaries). Finally, for the applicant, there is a greater inconvenience cost, with additional time required for travel to attend the proctored test environment, in addition to the time needed to actually do the test. This may also include the cost of taking time away from work, or from the family, to attend a testing session.

10.3.2.2 Computer-Based Testing

The use of computer-based testing reduces the costs associated with printing and supply of numerous copies of hard-copy forms, and it may also reduce some of the training requirements for scoring (i.e., if the system can score the test and produce a summary report itself). However, there are additional cost-related considerations for computer-based tests that might not be present for pencil-and-paper testing. Note that while computer-based tests may be administered in either a proctored or an unproctored environment, the focus in this section is on the costs of proctored computer-based testing.

For the test developer, the costs of developing a computer-based test from scratch (as described under *Common Cost-related* Factors) will still apply; alternatively, converting an existing pencil-and-paper test into a computerised form will also incur certain additional costs. These additional expenses include costs associated with test equivalency research, which is necessary to ensure the two test formats have the same utility and construct validity, and potentially the development of new computer-specific test norms. They might also include costs associated with paying programmers to convert existing tests (including consideration of such factors as screen size, computer speed/capability, and network reliability), and costs arising from paying graphic designers to ensure the test's graphical interface is both acceptable to candidates and does not interfere with the construct being tested.

With regards to test development costs, computer-based testing might also lead to increased costs due to the requirement for larger item pools, particularly when the test is adaptive. As noted previously, computer-based testing has enabled the development of new theories and models of psychometric testing, including CAT. As Fulcher (2000) points out, CAT requires a much larger item pool than is the case for traditional pencil-and-paper testing. The development and maintenance of this large item bank is time-consuming and costly for test publishers and researchers responsible for developing and maintaining such tests, meaning that the financial benefits arising from happy test takers and administrators might well be countered, at least in the shorter term.

Another model requiring a larger item pool is Linear-on-the-Fly Testing (LOFT). LOFT is essentially a compromise between traditional testing and CAT and works by ensuring each candidate is presented with a different but equivalent set of test items (based upon the item difficulty levels) drawn from a large item pool. In effect, this results in a test that can be seen as having a very large number of parallel forms – the more items in the item pool, the more parallel versions of the test that become available, reducing the likelihood of item over-exposure, and enhancing test security. Development and maintenance of larger item pools for CAT, LOFT and other computer-based (be it internet- or intranet-based) testing platforms clearly involves additional costs for test developers [9].

For the employer, the cost of maintaining a suitable testing environment (as required for proctored testing) still stand with computer-based testing, with some possible additional costs associated with ensuring the room is capable of supporting this form of testing. While there is no longer a cost associated with the printing of paper forms, there are costs associated with the purchase/hire and ongoing maintenance of computer hardware and software, such as workstations, monitors, printers, and data storage. There are also personnel costs associated with IT support, including security and data management.

10.3.2.3 Unproctored Testing

Unproctored, internet-based testing introduces still more additional costs for the test developer, employer, and applicants. For the test developer, the cost of developing larger item pools (as described above) applies, and there are additional costs associated with research and security. In terms of research, in addition to the previously identified maintenance and research costs, there are costs associated with ensuring test equivalency across both test forms (proctored and unproctored) and computer platforms. In terms of security, because of the risk of cheating in unproctored test environments (described in Chapter 5), there may be the need to develop enhanced security measures and protocols (e.g., keystroke monitoring and biometrics). This includes costs associated with verification testing.

Verification testing requires each applicant who has successfully completed the online, unproctored version of a selection test, to attempt the same test (or perhaps a shorter version of it) under standardised conditions (i.e., proctored) before continuing through the selection process. This process provides the employer with a final, reliable, and valid check of the candidate's credentials, by confirming – or otherwise – the test results achieved through unproctored online testing.

If the employer decides to maintain a 'test verification' process in addition to the unproctored internet-based testing, then all the costs associated with maintaining a proctored test centre will continue to apply, in addition to the costs identified above associated with a shift to unproctored testing. Retaining the requirement for proctored verification testing will clearly negate some of the advantages usually associated with unproctored internet-based testing (e.g., reduced staffing costs, convenience for applicants [8]); however, proctored verification testing is expected to be on a somewhat reduced scale, given that some applicants will already have been screened out at the initial online stage [7].

Note that a decision to utilise an unproctored testing model *without* verification testing may save money, at least in the short term, by removing the need to maintain and staff a separate, dedicated test centre. However, it also introduces additional risk into the selection testing process, in the form of reduced test security, less standardisation/control, and the like. These additional risks can lead to significant costs for the employer in the longer term, due to an increase in both Type 1 ('false positive' – someone being selected who should not have been, leading to increased risk of training failures and adjustment issues) and Type 2 errors ('false negative' – someone being rejected or self-selecting out based upon an invalid test score).

For the employer, there will be the need to develop a new policy framework specific to support an unproctored testing regime, including dealing with abnormal response patterns and other evidence of 'cheating', consideration of test waivers, and duty of care responsibilities. There are also additional costs associated with website and webpage design, and provision of real-time helpdesk support (covering both IT and test-based issues). This latter point leads to the requirement for helpdesk personnel to be available outside of core business hours to provide support to those completing the test late at night or early in the morning. Finally, costs associated with maintaining a proctored, computer-based, or pencil-and-paper test still apply (even if the employer does not decide to use verification testing, as described above), as a secondary test option needs to be available for applicants who either do not have internet access at home or prefer to undergo testing via a different platform.

10.3.2.4 Applicant Costs

Are there additional costs for the applicant as well? In an ideal world, an unproctored test would be able to be run on any device that an applicant might own. In the real world, however, this is simply not the case – the huge range of device types and sizes available to consumers means that there will inevitably be some devices

(and networks) upon which a given test works better than it does on other devices. In addition, even if the test works perfectly upon any device, the applicant may need to purchase new hardware and/or software, such as a webcam, an application, or a different web browser, according to the protocols put in place by the employer. Other factors that might need to be considered include ensuring that download/upload speeds are sufficient to enable the test to run smoothly without buffering, and that the device has sufficient data capacity to run the test. The issue of testing on mobile devices is explored in greater detail in Chapter 11 of this report.

10.4 SUMMARY AND CONCLUSION

In this chapter, we have compared different models of testing environments (proctored pencil-and-paper, proctored computerised, and unproctored) in terms of key cost-benefit considerations. This chapter acknowledges the many potential benefits that await those employers who choose to adopt psychological testing as part of their recruitment and selection strategy. However, it also identifies that there are also costs associated with the various testing models, and that the cost-benefit trade-off can vary considerably. Each testing model has costs for the test developer, the employer, and the applicant; some of these costs are common across all models, while others are specific to some approaches but not others. Different testing models also have different risks associated with them and mitigating these risks can incur additional costs. The next chapter addresses some of the risks and risk mitigation strategies associated with proctored and unproctored testing environments (including the direct and indirect costs associated with the risks), along with some of the other factors that warrant consideration when deciding which model of testing to use.

10.5 REFERENCES

- [1] Bauer, T.N., Truxillo, D.M., Paronto, M.E., Weekley, J.A., and Campion, M.A. (2004). Applicant reactions to different selection technology: Face-to-face, interactive voice response, and computer-assisted telephone screening interviews. *International Journal of Selection and Assessment*, 12, 135-148.
- [2] Carless, S.A. (2009). Psychological testing for selection purposes: A guide to evidence-based practice for human resource professionals. *The International Journal of Human Resource Management*, 20, 2517-2532, doi.org/10.1080/09585190903363821.
- [3] Hinton, M. (2005). Review of the use of internet-based testing in recruitment and selection. Research Report 03/2005: Psychology Research & Technology Group, Canberra.
- [4] Hinton, M. and Swann, J. (2013). Online testing: challenges and promise. In P.J. Johnston and K. Farley (Eds.), *Military human resource issues: A multinational view*. Kingston, Ontario: Canadian Defence Academy Press, 97-111.
- [5] International Test Commission (2005). International guidelines on computer-based and internet-delivered testing. https://www.intestcom.org/files/guideline_computer_based_testing.pdf (Accessed 10 Jan 20).
- [6] McCarthy, J.M., Bauer, T.N., Truxillo, D.M., Anderson, N.R., Costa, A.C., and Ahmed, S.M. (2017). Applicant perspectives during selection: A review addressing “so what?” “what’s new?,” and “where to next?”. *Journal of Management*, 43, 1693-172.
- [7] Mooney, J. (2002). Pre-employment testing on the Internet: Put candidates a click away and hire at modem speed. *Public Personnel Management*, 31(1), 41-52.

- [8] Naglieri, J.A., Drasgow, F., Schmit, M., Handler, L., Prifitera, A., Margolis, A., and Velasquez, R. (2004). Psychological testing on the internet: new problems, old issues. *American Psychologist*, 59(3), 150-162.
- [9] Parhizgar, S. (2012). Testing and technology: Past, present and future. *Theory and Practice in Language Studies*, 2(1), 174-178.
- [10] Ree, M.J., Earles, J.A., and Teachout, M.S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology*, 79(4), 518-524. doi.org/10.1037/0021-9010.79.4.518.

Chapter 11 – INTERNET-BASED TESTING ON MOBILE DEVICES

Geoff Gallas
Department of Defence
AUSTRALIA

Mobile devices are everywhere! As a society, we use mobile devices for everything that we do, from shopping lists, to monitoring our children's whereabouts, to setting our thermostats remotely; mobile devices have become our constant relied upon tool for navigating our world. This begs the question, why would seeking a job be any different? ([17], p.4)

11.1 INTRODUCTION

Over the last few decades, the world has seen incredible advances in the field of information and communications technology. These advances, particularly in the area of enhanced processing speed and power, have enabled the development of new theories and techniques that have in turn had an enormous influence on psychometrics and test technologies, such as Item Response Theory (IRT) and Computer Adaptive Testing (CAT). The emergence in the early 1990s of the Internet and the World Wide Web [6], and its evolution into a viable (and affordable) commercial and personal computing platform, essentially enabled the next step in the evolution of tests and testing technology.

Recruiting agencies soon realised that there were likely to be certain benefits, for both employers and potential job applicants, in allowing applicants to undertake certain tasks (e.g., filling in an application form, submitting educational documentation) online, in a familiar environment, using equipment with which they felt comfortable. This same thinking started to extend into the testing arena. According to the Australian Psychological Society (APS) [2], survey results suggest that an increasing number of HR professionals claim to have used online rather than paper-and-pencil testing as part of their recruitment campaigns [7], [8] as far back as the start of the decade, with a significant proportion of those assessments being conducted remotely. This number will only have increased in the years since that particular survey was conducted.

So, what is the next 'evolution in testing affairs'? The popularity of the smartphone since that technology burst onto the scene post-2007 makes the mobile device an obvious favourite. Shalfrooshan [22] claims that in 2017, approximately two-thirds of all internet-based traffic in the United States originated from mobile phones, and that this figure was only expected to grow. With mobile devices becoming ever more ubiquitous in all aspects of our lives, it's only natural that job seekers / applicants will come to expect to be able to apply for jobs, and undergo some form of assessment for those jobs, using these devices.

In this chapter I will identify both the potential benefits most often cited by proponents of mobile-based testing, and the main costs that a decision to progress down this pathway is likely to incur, for both the organisation and the individual test taker, when developing and implementing this technology.

11.2 POTENTIAL BENEFITS OF TESTING ON MOBILE DEVICES

As identified by the APS [2], use of mobile devices for testing is growing. After all, testing on mobile devices offers a number of potential advantages to both the employer and the candidate (note that the potential advantages of internet-based testing in general are addressed in greater depth elsewhere in this report):

- Mobile devices make light, portable test delivery platforms.
- Mobile devices are usually always switched on, so there are minimum delays caused by the system needing to ‘boot up.’
- Mobile testing is convenient for both test takers, who can take the assessment wherever and whenever they want, and test administrators, who can allow tests to be administered 24 hours a day with little if any human involvement.
- Mobile testing might be seen as modern, progressive and ‘high tech’, and hence serve as an attractant for various segments of society, or people with particular interests.
- Mobile testing can (but might not always) result in significant cost savings for the employer, due to there no longer being a need to maintain a physical test administration centre.

As mobile devices become more and more capable, the kinds of tests that might be administered via this medium, and the types of measurements that might become possible (e.g., response/reaction times), will only increase.

Lawrence and Kinney [17] also note that there is some evidence that different minority groups might be more likely to access the internet via their smart phone than from a fixed desktop Personal Computer (PC). This offers the potential for far greater reach into the community than might otherwise be the case with traditional testing models, which may in turn contribute to greater diversity within the workforce.

11.3 POTENTIAL COSTS OF TESTING ON MOBILE DEVICES

11.3.1 Costs for the Organisation

General resource implications arising from the question of whether to pursue an internet-based online testing regime as part of a military organisation’s recruitment and selection program are addressed in Chapter 9 of this report. This section explores the additional costs that may need to be taken account by a military organisation considering making tests available on mobile devices such as tablets and smartphones.

11.3.1.1 Test Development Costs

Tests designed for use on a PC or laptop might perform less well on a mobile device; similarly, a test designed specifically for use on an Apple product (for example) might not perform to the same specifications when delivered via a device designed for a competing operating system, such as Google or Microsoft. As previously identified, assessments need to measure candidate characteristics consistently across devices, to ensure all candidates have access to a test medium that a) they are comfortable with, and b) will afford them the best opportunity to perform to their optimum level.

The mobile and tablet market in particular is highly fragmented, as evidenced by the vast number of manufacturers and models available to consumers, not to mention the various competing operating systems that underpin those devices. Variables that would need to be considered here when proposing a move to mobile testing might include device manufacturer, market share and popularity; different operating systems and different versions of those operating systems; device screen sizes; viewports and screen resolution; upcoming developments and new devices being released; and launch year [4].

Developing and testing different versions of a test on the maximum number of mobile device options available would necessarily involve a huge investment in time, effort, and money. As a result, it would be unrealistic (if not impossible) for any military organisation to design, monitor and maintain different versions of even a small number of core selection tests to suit all possible test delivery options. A more realistic and affordable option might be for the organisation to develop a small number of tests designed specifically for use on the most common devices (e.g., PC/laptop – Microsoft and MacOS; mobile device – Android and IOS), and to specify to potential applicants that these are the only acceptable options. Note that Lawrence and Kinney [17] actually warn employers that simply instructing candidates not to use certain mobile devices is not sufficient; rather, your technology must actually prevent the assessment from opening if the applicant attempts to access the assessment on a non-approved device. Another more affordable option might be to limit the types of assessments made available via a mobile device (e.g., initial application form, biodata form; personality test), and maintain centralised test administration centres for more high stakes cognitive testing requirements.

11.3.1.2 Test Maintenance and Version Control Costs

Of course, issues to do with test development do not end once the first version of a test is launched. Just as traditional pencil-and-paper tests require psychologists and psychometricians to constantly evaluate the internal workings of each test and validate test results against organisational outcomes, to ensure they remain fit for purpose, the same requirements apply for computer- and internet-based tests. With respect to mobile devices, this would require the test publisher to continually monitor developments in the mobile Information and Communications Technology (ICT) industry, and update tests in response to ongoing changes and upgrades to devices and their operating systems, to ensure optimal candidate performance (i.e., tests no longer capable of running on outdated operating systems; patches designed to counter system ‘glitches’ and ‘bugs’), and ensure the tests themselves remain secure [4].

11.3.1.3 Other Indirect Risks and Associated Costs

As discussed in Chapter 10, there are a number of indirect costs or risks associated with implementing any psychometric testing program, and these risks become greater the more removed the test administrator becomes from the test taker. In the case of mobile testing, there is a greater risk of a) an ‘incident’ (e.g., the test fails to load); b) which is not noted by the test administrator; and hence c) is not contained, controlled, or resolved in an appropriate timeframe. Similarly, the lack of an on-site test administrator in this sort of circumstance also limits the ability of the employer to enter into effective ‘damage control’ mode by directly controlling the situation on the ground in real time. These factors, separately or together, might well contribute to increased incidence of subsequent appeals, complaints, and grievance action.

All the various recruitment and selection models and combinations of tools, techniques and processes present different risks and involve different costs. An employer opting to allow applicants to use a mobile assessment platform for more low-stake activities (e.g., completing and submitting an application form) would then be responsible for the costs associated with establishing and maintaining a physical test centre for applicants to attend to complete the ‘high stakes’ cognitive testing component of the process. On the other hand, an employer choosing an entirely un-proctored, mobile-based selection process would have no one else to blame when candidate feedback suggests the organisation is seen by applicants as remote and impersonal.

11.3.2 Costs for the Individual

Resource costs associated with a move to mobile testing are not restricted to the employer organisation alone. The indirect risks and costs of poor system design and implementation identified above might look to be all

one-sided (i.e., against the organisation), but each risk for the organisation also means a potential reciprocal risk/cost for the individual applicant. For example, lack of responsiveness and poor internet download speeds can lead to applicants feeling frustrated and angry due to the perceived waste of time, effort, and money (i.e., wasted data). This might in turn contribute to self-selection out of the recruitment process, and disinclination to pursue what might otherwise have turned out to be a fantastic career – or, perhaps worse, failing to meet the required standard on a particular selection test, and hence failing to be considered for enlistment, due to poor system performance outside the applicant’s control. This would only be exacerbated in a mobile-only-based assessment situation, in which online test results are considered in isolation from any other evidence, and the applicant does not have the opportunity to challenge the test result with selectors in person.

11.4 RISK AND RISK MITIGATION

11.4.1 Potential Disadvantages and Issues Associated with Testing on Mobile Devices

Note that issues common to all internet-based testing – including screen sizes; device capability; network reliability/stability; possibility of cheating (e.g., having other people help answer the test items; looking answers up on other devices); test security issues (e.g., items being compromised or stolen) are discussed elsewhere in this report.

11.4.1.1 The Very Nature of Being ‘Mobile’

One particular concern associated with conducting psychometric testing on a mobile device is that, by their very nature, both the test taker and the testing platform/medium are ‘mobile’. Preuss [21] notes that while taking assessments on the move might be convenient, the situation in which one chooses to attempt the test might also prove distracting, which can compromise performance. Attempting to complete a complex test of general cognitive ability in an internet café or an airport lounge, at a coffee shop or while waiting for a bus, is not likely to prove conducive to optimal test performance [2]. Chang and colleagues [5] also noted that not only were candidates who use mobile devices to complete assessments more likely to encounter distractions during the assessment process, those distractions also appeared to have a more substantial negative impact on scores for applicants using mobile devices, compared to those undertaking testing on a PC. However, more recent work by Traylor and colleagues [24] found that performance on a test of general cognitive ability did not appear to be affected by the testing environment, albeit in a laboratory rather than an operational environment.

The advent of test delivery on mobile devices clearly increases the likelihood of greater variability being present in the testing environment. Poor internet connectivity and other sources of variability amongst internet service providers (e.g., download speeds, security applications designed to protect the device) will also have an adverse effect upon the testing environment and the quality of the candidate experience. Any psychologist working in assessment would agree that one of the basic tenets of testing is that no candidate should suffer any disadvantage as a result of the medium within or upon which they are tested. Every potential job applicant has a right to expect that an organisation’s recruitment and selection policies, procedures and standards are fair regardless of culture, age, disabilities or gender [21]; they also therefore have a right to expect that the same degree of ‘fairness’ will apply whether they undergo testing on a paper-based form, a PC, or a mobile device.

11.4.1.2 Fairness and Test Equivalence

As noted previously, several studies have suggested that members of certain minority groups are more likely to use a mobile device to access the internet than a desktop PC or laptop. McClure-Johnson and Boyce [18] showed

that in a large American sample, women, African-American and Hispanic, and younger applicants were more likely to use a mobile device for testing than other applicants. This was supported by Smith [23], who noted that in the US, certain demographic groups, particularly those with African-American or Latino heritage, were more likely to rely on their device as their only source of internet access. Still, other research has suggested that candidates from Asia, as compared to the Americas and Europe/Africa, are more likely to express a preference for undertaking employment assessments on mobile devices [2], [8].

This trend has perhaps more to do with socioeconomic and cultural factors than to do with race. For example, Shalfrooshan [22] claims that figures published by the US Office for National Statistics (ONS) have suggested that there remains a divide between different socioeconomic groups regarding the technology they have access to, rather than just the technology they prefer. Whatever the reason, as identified previously, this means that a mobile-friendly recruiting and selection process has the potential to gain greater access to minority groups, and hence improve their representation in the candidate pool.

Of course, this positive outcome depends very much upon the tests themselves being fair and performing equivalently for all applicants, regardless of education, culture, age, disabilities, or gender [21]. If the tests are not in fact fair, the result might well be reduced rather than increased diversity. As Lawrence and Kinney [17] point out, employers need to ensure that adverse impact does not occur as a result of their recruitment, selection and/or employment policies and practices. For example, if more applicants from a particular minority group are known to utilise mobile devices more frequently than applicants from other groups, and those minority applicants also perform less well on those tests (or rather, if the version of a test delivered via mobile device does not allow them to perform as well as they might on the same test delivered via PC), then adverse impact increases, undermining the potentially positive effects (i.e., increased reach) of mobile testing.

This suggests that militaries considering utilising mobile devices for employment-related testing need to have a very sound understanding of their national demographic and how that might influence access to and use of mobile devices (e.g., possible regional and/or cultural attitudes towards technology; factors such as socioeconomic status). They also need to ensure that different versions of tests deployed across various platforms have been demonstrated to be fair (on racial and gender grounds as a minimum) and not result in adverse impact for any identified minority group.

11.4.1.3 Equity of Access – People with a Disability

Questions regarding test equivalence can also be extended to include issues to do with equality of access, especially for people with a disability. As labour markets tighten and recruiting targets become harder to fill, militaries will start looking to new, previously untapped (or, at best, under-tapped) sources to fill certain roles – this can and will include people with a recognised disability. As stated previously, no candidate should suffer any disadvantage as a result of a test being offered on a particular platform, be that a desktop PC or a mobile device [21]. International Test Commission (ITC) [14] guidelines require that organisations ‘consider reasonable adjustments to the technical features of the test for candidates with disabilities’, a requirement that conceivably not only has implications for the physical hardware and other technology used to deliver the test, but also for the design and content of the very test itself. The word ‘reasonable’ is left open to interpretation – moreover, the extent to which such ‘adjustments’ are possible, or perhaps even desirable, will become a matter of policy for the military organisation in question, and might be expected to vary from country to country.

11.4.1.4 Device Type

Some researchers have suggested that the type of device used for an assessment might impact upon outcomes differently, depending upon the type of test being administered. Many studies (e.g., Refs. [1], [13], [15]), amongst others, all showed that people undertaking cognitive testing (e.g., tests of General Mental Ability) and tests of response speed / reaction times on mobile devices tended to perform more poorly than people doing the same tests on more traditional PC or laptop-type platforms. On the other hand, performance on non-cognitive assessment tools, such as personality tests, biodata forms and text-based Situational Judgement Tests, appeared to be the least affected by device type. Huff [11] also reported higher error rates when people were asked to complete tests that had not been specifically designed (or optimised) for use on mobile devices.

Brown and Grossenbacher [3], however, found that neither device type nor screen size had any practical impact upon test performance on a test of General Mental Ability, especially when test content and item presentation are optimised for use on mobile devices (e.g., changing a response format to remove the need to scroll to see additional response options on smaller screens). Similarly, Neale and Wang [19] also found that neither device type nor screen size had a significant main effect on test performance. They did, however, observe a significant interaction effect between *device preference* and *device type*, and concluded that it was likely that applicants would tend to perform better on tests (including cognitive ability tests) if they were able to undertake those tests on the device of their choice (be it PC, laptop, tablet, or smart phone). In the educational domain, some researchers (e.g., Refs. [12], [25] as cited in Ref. [20]) have also shown that it is possible to develop web-based tests and tools that are both valid and reliable for the context in which they are used.

Traditionally, as new technologies have evolved, the most common (also the most logical, cost efficient, etc.) approach has been to ‘tweak’ pre-existing proven tests and administer them via the new platform. However, if the goal of testing is to provide the candidate with the opportunity to perform to the best of his/her ability whatever the test delivery platform, tests must be developed or adapted specifically to suit the device in question. This needs to include not just test content and design but overall usability from a human factors perspective as well [2], [14]. Shalfrooshan [22] has suggested a change to the traditional test adaptation paradigm, via the concept of ‘mobile-first design’ (emphasis added):

*As the name suggests, **mobile-first design** is the process of designing a website or piece of software for mobiles or small-screened devices **first**, then working up to large-screened devices. This isn’t a new phenomenon and has been around for a few years but now that [mobile devices] are the primary device used for browsing the web, this has specifically impacted recruitment ... Simply making a test compatible with mobiles is not enough; the format, what it is measuring, and the **usability** needed to be readjusted and rebuilt to **ensure the candidate experience is optimal**.*

Of course, on a more prosaic level, the concept of usability must also encompass the physical qualities or status of the device in question. Just as a broken hand might detract from an applicant’s performance on a paper-and-pencil test, so too might a cracked or broken screen on a tablet or smartphone detract from test performance on that particular device. Test administrators need to keep these kinds of issues in mind when developing test instructions and interpretation guides, whatever the medium they are working in.

11.4.1.5 Multi-Stage Assessment – Different Tools for Different Jobs?

Arthur and colleagues [1] suggest that the type of device upon which a test is delivered can affect performance on different types of tests (e.g., cognitive vs non-cognitive tests) in different ways. For example, it was noted that people completing a cognitive ability test on mobile devices tended to perform significantly worse than people completing the same tests on non-mobile devices; this impact was, however, not noted in the case of

non-cognitive personality tests. If this finding can be more widely and reliably validated, it suggests that that different delivery platforms might be best utilised for different parts, and at different stages, of the assessment process.

For example, a text-based application form (un-timed), a general health questionnaire, and perhaps a personality test of some description might be provided to the potential applicant early in the recruitment process, for un-proctored completion on their device of choice in their own home at a time that suits them. For high stakes assessment including cognitive ability testing, however, an employer could specify (or even ensure technologically) that an applicant can only attempt the assessment using a particular type of device (e.g., tablet, laptop or PC, but not a smart phone), or require the applicant to attend a designated testing centre to undertake that part of the assessment [21]. Such a multi-faceted recruitment and selection process, if designed properly, would enable the employer to leverage most of the advantages of remote internet-based testing (e.g., convenience, increased access), whilst also ensuring the advantages associated with a more controlled, standardised approach (encouraging optimal candidate performance) are also recognised.

11.4.1.6 Applicant Reactions

How good the tests themselves are, how reliable the internet connection is, and how controlled the testing environment all matter little if applicants do not respond well to the assessment experience. While a well-designed and implemented testing program might be considered a reputational builder or enhancer for an organisation, a poor applicant experience, whether due to technical system issues, poor test design, even poor communications with the employer, will act as a deterrent.

But do applicants react differently to testing on mobile devices than they do when the same tests are delivered via PC? King and colleagues [15] suggest that applicants do tend to react differently to tests delivered via a mobile device compared to a PC, with applicants tested on a PC reporting significantly more positive reactions to testing than those tested on a mobile device. Indeed, 92% of candidates surveyed in one particular study stated that they preferred to use a PC or laptop for ‘high stakes’ testing rather than a mobile device [9].

Shalfrooshan [22] supports this statement, claiming that just because smartphone usage is becoming ubiquitous in general life, with usage growing significantly, this may not be reflected in a candidate’s preferences regarding which test platform they would prefer. Citing internal (unpublished) research with university graduates, Shalfrooshan claims that 49.8% of subjects indicated that they would be most likely to use a laptop to complete online assessments, while just 30.4% indicated they would prefer to use a smartphone, suggesting the result could likely be put down to graduates basically being ‘intelligent’ (although perhaps ‘knowledgeable’ or ‘self-aware’ might be a better term) enough to realise that doing tests in more controlled environment would likely be in their own best interests. Alternatively, this sort of result might also arise from some applicants being more committed to the employer and/or the process than others, and hence more likely to wish to optimise their chances of performing well.

As noted previously, applicants’ test performance on mobile devices depends at least in part upon whether the test has been designed specifically for use on that device [2], [3] Similarly, applicant reactions to undertaking testing on a mobile device also appear to depend upon whether the test was designed specifically for that device [16], [17], or simply converted from some alternative platform. It is suggested that the primary underpinning factor here might well be a simple issue of functionality [10] – as systems and applications improve in general, and peoples’ expectations lift in response, any system that does not work smoothly will not be put up with by people in our increasingly rushed and time-poor society.

11.4.2 Managing Risk in a Mobile Device-Based Selection Context

It is generally accepted by psychologists that the best predictor of future work and training performance is past, or better still current performance. However, work experience of this sort is expensive and, in organisations like the military, potentially hazardous, given that unlike civilian employers the military does not usually expect its recruits (especially in Combat Arms roles) to come in pre-trained and job-ready. As a result, various cheaper and simpler proxy measures of work potential are generally employed in place of work experience. These proxy measures cover the entire gambit of selection procedures, from work sample tests and assessment centres, through psychometric testing and interviews, to simple resume checking. These approaches tend to differ in terms of ease of application, cost, and validity; and as is usually the case, the best (i.e., most valid) solution will also generally prove to be the most resource intensive and expensive.

With this in mind, what must be remembered when weighing the pros and cons of both internet-based testing and mobile testing is, essentially, that all psychometric testing in the personnel selection arena constitutes a risk – the better the risk management strategies employed (e.g., good test validity; consistency across platforms), the better the quality of the employment prediction made, and the better the outcome for both employer and employee.

11.5 SUMMARY AND CONCLUSION

This chapter looks at the increasing use of mobile devices as a means of delivering internet-based selection tests to applicants in a place and at a time of their own choosing. The advantages associated with mobile device-based assessment probably make this approach very attractive to recruitment agencies and organisations having to fight not just for talent in the labour market but also for internal resources. However, as this chapter demonstrates, for every perceived advantage or benefit that might come from conducting selection testing on a mobile device, there is likely to be a set of reciprocal costs associated with realising those benefits. Moreover, while increasing sophistication within a selection testing program might bring certain benefits to both the organisation and the individual applicant, these benefits must at all times be weighed against the potential risks resulting from reduced levels of organisational control over the test taking experience, and the longer-term costs associated with those risks.

Accepting that different aspects of the recruitment and selection process might benefit from different approaches and different platforms and adopting a multi-faceted approach to military recruitment and selection, might for some countries prove a cost-effective way of managing the selection assessment process in a way that best balances the ‘risk vs reward’ equation in their national context. Offering job applicants the opportunity to undertake at least *part* of the military recruitment and selection process (including testing) on their own preferred device, be that mobile or fixed, at a time and place of their choosing, is likely to come to be considered an inevitable part of military recruitment. After all, as Preuss [21] points out:

... your candidates are not only likely to have a smartphone and a tablet, they're also likely to feel comfortable using these devices to conduct a range of different tasks, including searching for a new job. Offering your assessments [or at least a component thereof] via these devices is therefore the next logical step.

Ultimately, senior military decision makers must carefully weigh up all the potential risks and rewards offered by the various combinations and options available to them, in order to make a fully informed decision that will (hopefully) best enable the organisation to meet the requirements of their mission.

Decision makers will need to consider very carefully a number of factors, ranging from psychometric, technological and financial factors, cultural mores and societal norms, and organisational appetite for and tolerance of risk, in determining whether their particular military organisation wishes to proceed down this pathway.

11.6 REFERENCES

- [1] Arthur, W., Doverspike, D., Muñoz, G.J., Taylor, J.E., and Carr, A.E. (2014). The use of mobile devices in high-stakes remotely delivered assessments and testing. *International Journal of Selection & Assessment*, 22(2), 113-123.
- [2] Australian Psychological Society Tests and Testing Expert Group (2014). *Online psychological testing*. Australian Psychological Society, Melbourne, Australia.
- [3] Brown, M.I. and Grossenbacher, M.A. (2017). Can you test me now? Equivalence of GMA tests on mobile and non-mobile devices. *International Journal of Selection & Assessment*, 25, 61-71.
- [4] Browserstack (2019). Test on the right mobile devices. <https://www.browserstack.com/test-on-the-right-mobile-devices> (Accessed 11 Sep 19).
- [5] Chang, L.C., Lawrence, A.D., Kinney, T.B. and O'Connell, M.S. (2016). What was that? Investigating distractions on test performance and reactions. Paper presented at the 31st annual conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.
- [6] Couldry, N. (2012). *Media, society, world: Social theory and digital media practice*. London: Polity Press.
- [7] Fallaw, S.S. and Kantrowitz, T.M. (2011). 2011 global assessment trends report. SHL PreVisor.
- [8] Fallaw, S.S., Kantrowitz, T.M., and Dawson, C.R. (2012). 2012 Global assessment trends report. SHL PreVisor.
- [9] Fursman, P.M., and Tuzinski, K.A. (2015). Reactions to mobile testing from the perspective of job applicants. Paper presented at the 30th annual conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.
- [10] Gutierrez, S.L., and Meyer, J.M. (2013). Assessments on the go: Applicant reactions to mobile testing. In N.A. Morelli (Chair), *Mobile devices in talent assessment: Where are we now?* Symposium at the 28th Annual Conference of the Society for Industrial and Organizational Psychology, Houston, TX.
- [11] Huff, K.C. (2015). The comparison of mobile devices to computers for web-based assessments. *Computers in Human Behavior*, 49, 208-212.
- [12] Huang, Y.M., Lin, Y.T., and Cheng, S.C. (2009). An adaptive testing system for supporting versatile educational assessment. *Computers and Education*, 52, 53-67.
- [13] Illingworth, A., Morelli, N., Scott, J. and Boyd, S. (2014). Internet-based, unproctored assessments on mobile and non-mobile devices: Usage, measurement equivalence, and outcomes. *Journal of Business and Psychology*, 30(2), 325-343.

- [14] International Test Commission (2005). International guidelines on computer-based and internet-delivered testing. https://www.intestcom.org/files/guideline_computer_based_testing.pdf (Accessed 10 Jan 20).
- [15] King, D., Ryan, A.M., Kantrowitz, T., Grelle, D. and Dainis, A. (2015). Mobile internet testing: An analysis of equivalence, individual differences, and reactions. *International Journal of Selection and Assessment*, 23(4), 382-394.
- [16] Kinney, T.B., Lawrence, A. D., and Chang, L. (2014). Understanding the mobile candidate experience: reactions across device and industry. In T. Kantrowitz and C. Reddock (Chairs), *Shaping the future of mobile assessment: Research and practice update*. Symposium at the 29th Annual Conference of the Society for Industrial & Organizational Psychology, Honolulu, HI.
- [17] Lawrence, A.D., and Kinney, T.B. (2017). *Mobile devices and selection*. SIOP White Paper Series; Society for Industrial and Organizational Psychology, Inc.
- [18] McClure-Johnson, T.K. and Boyce, A.S. (2015). Selection testing: An updated look at trends in mobile device usage. In Neil Morelli (Chair), *Mobile devices in talent assessment: The next chapter*. Symposium at the 30th Annual Conference of the Society for Industrial and Organizational Psychology, Philadelphia, PA.
- [19] Neale, M. and Wang, S. (2019). Game-based cognitive ability testing: Impact of mobile and desktop device usage on scores. Paper presented at the 13th APS Industrial & Organisational Psychology Conference, Adelaide, Australia.
- [20] Parhizgar, S. (2012). Testing and technology: Past, present and future. *Theory and Practice in Language Studies*, 2(1), 174-178.
- [21] Preuss, A. (2016). Should you offer your assessments on mobile devices? <https://www.hrzone.com/talent/development/should-you-offer-your-assessments-on-mobile-devices> (Accessed 09 Jan 20).
- [22] Shalfrooshan, A. (2018). How a mobile-first assessment process impacts recruitment. <https://blog.psonline.com/talent/blog/how-a-mobile-first-assessment-process-impacts-recruitment> (Accessed 09 Jan 20).
- [23] Smith (2015). The smartphone difference. Pew Research Centre. <http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/> (Accessed 09 January 2020).
- [24] Traylor, Z., Hagen, E., Williams, A. and Arthur, W. (2021). The testing environment as an explanation for unproctored internet-based testing device-type effects. *International Journal of Selection and Assessment*, 29, 65-80.
- [25] Triantafillou, E., Georgiadou, E. and Economides, A.A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers and Education*, 50, 1319-1330.

Chapter 12 – ONLINE TESTING: SUMMARY AND CONSIDERATIONS

Elliot Loh

Department of National Defence
CANADA

12.1 SUMMARY

Chapters 6 through 11 of this report explored both the advantages and issues that must be considered when implementing online personnel assessment systems for military organisations. Chapter 6 identified that properly implemented digital testing systems can provide an efficient approach to the assessment of applicants for military occupations. The cost of managing paper-based tests is eliminated, changes to test items can be made quickly and universally, and digital testing platforms provide the opportunity to present more complex and advanced assessments. Chapter 7 noted that the ubiquity of the internet and its pervasive use by younger adults who are typically the target of military recruiting initiatives all but demand that our applicant assessment systems must be implemented online to the fullest extent possible. Moreover, as militaries often compete with private sector firms, the implementation of online application and assessment systems is becoming essential in order to retain the image of a modern employer [1].

Moving assessment systems online brings with it advantages of convenient access for most applicants and reduction in the need for physical testing facilities and test administration staff. However, there is a trade-off; with online testing there is a significant reduction in test content security, an increase in the risk for cheating, and authentication of a test-taker's identity is more difficult. These concerns were shared by many participating nations, as described in Chapter 8, and Chapter 9 detailed approaches to the identification and mitigation of such risks. Beyond these issues, the development of an online testing system to support military personnel selection objectives requires the implementation of a range of new resources and processes which are not needed when using traditional paper-based tests. The costs of these along with the benefits they provide are detailed in Chapter 10. As was noted in Chapter 8, most nations are in the early stages of the implementation of online assessment systems; however, advanced technologies such as smartphones and tablets enjoy ubiquitous use. Accordingly, it is reasonable to assume that there is an emerging expectation from potential applicants to be able to use such technologies for assessment purposes in the not-so-distant future. Chapter 11 provides details on the challenges with implementation of smartphone technologies.

12.2 CONSIDERATIONS

The development and implementation of an online applicant assessment system requires a consideration of the factors noted in the preceding chapters. These considerations may be grouped into four interrelated domains that are summarised in Figure 12-1. Details of the considerations are provided below.

12.2.1 Online Software Implementation

Moving to an online implementation of an assessment system requires the consideration of a number of issues beyond that of a simple digital implementation of an applicant assessment. Most importantly, the capability to develop and maintain a moderately sophisticated website, and possibly, the support of mobile implementations,

ONLINE TESTING: SUMMARY AND CONSIDERATIONS

must be in place. Resources for this must either be present or will need to be acquired. The website must be able to support the security requirements such as user authentication, maintaining a secure repository of assessment items, as well as securely storing each test-taker's personal information and their responses to the assessments administered. In addition, the provision of online assessments can result in higher volume of website or mobile site traffic which the system must be able to support. As the online implementation of assessments will result in their presentation on a range of devices with different operating systems, screen sizes and user input approaches (e.g., mouse, keyboards, touch-screens), it is important to enable quality assurance strategies to ensure the assessments are displayed properly and function properly on all devices authorised for test administration. As completion of the assessment may be considered higher-stakes and time-dependent, the level of on-going test-taker support though helpdesks may need to be reviewed.

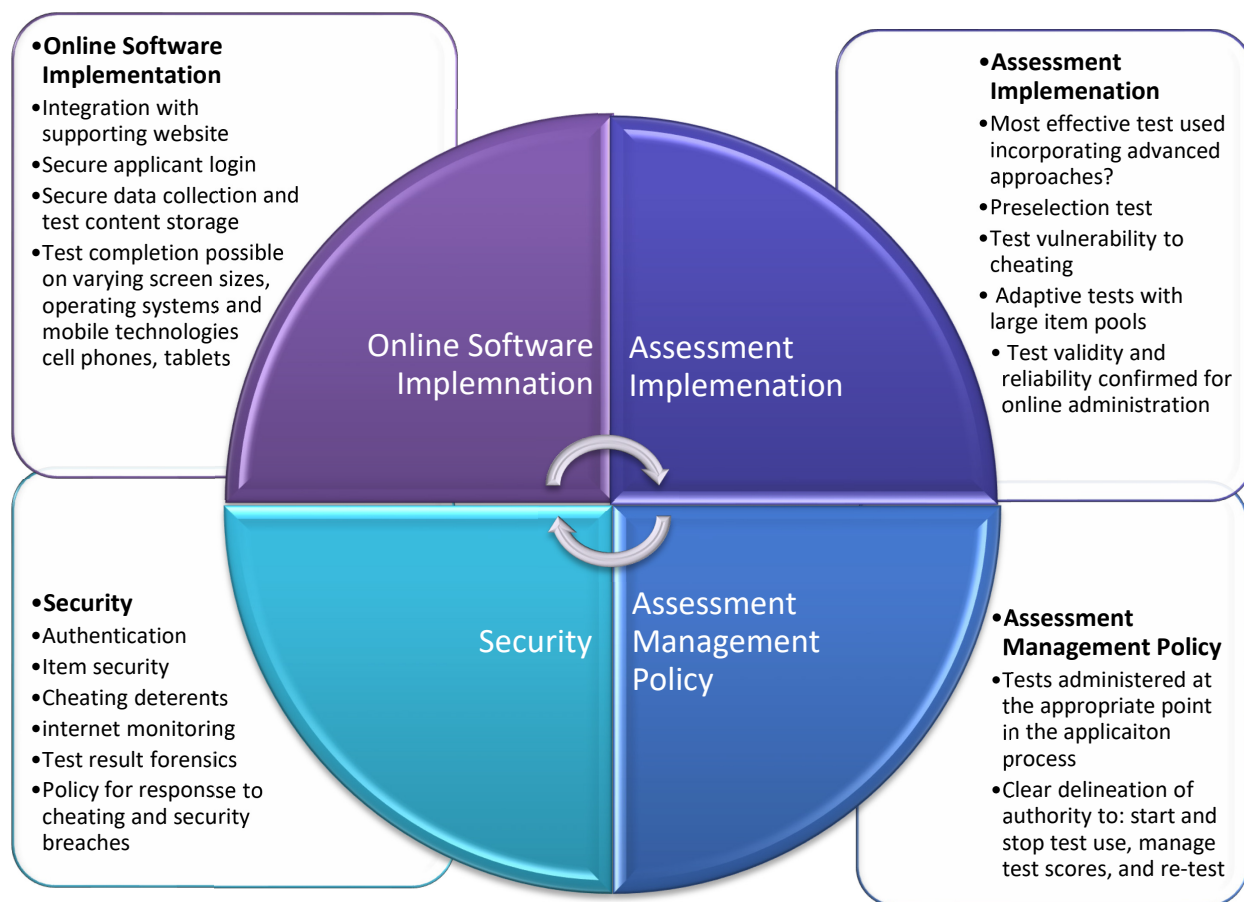


Figure 12-1: Considerations When Implementing Online Assessments.

Online assessment functionality will likely be integrated with the organisation's website and as such the look and feel of assessment tools and the work-flow for the completion of the assessments should fit seamlessly with the existing site. To support online implementation, a high level of communication between the test development group, the operational test management group, and the IT professional managing the online technology is essential. Each group should develop a good understanding of the other groups' priorities and requirements.

12.2.2 Security

The security considerations related to the implementation of online assessments expand beyond the typical concerns that might be covered through the development of other websites. For example, user authentication and cheating prevention are often critical features. An online webcam-based proctoring capability allows for easier authentication of the test-takers' identity, and may reduce likelihood of test-takers cheating, but the systems are not infallible for the prevention of cheating or identity spoofing. The security of the test content is also paramount. Additional features may be added to the online implementation to ensure the assessments are not compromised by enabling digital monitoring algorithms that track sudden changes in test scores. Websites known for disseminating restricted information on assessments should be monitored for test content security breaches.

While it is important to enable appropriate security measures, it is equally important to have in place policies and plans to quickly deal with breaches of security. Responses to situations such as the discovery that a test has been published to a social media site, or a webcam proctor suspects, but is not entirely sure, that a test-taker is cheating, must be determined in advance. It is also important that test-takers are aware of the security precautions in place in order to establish a sense of fairness for all applicants. Applicants who complete their assessment honestly must not perceive that they are at a disadvantage to those who potentially could have cheated, regardless of whether others are cheating or not.

12.2.3 Assessment Management Policy

Policies related to assessment management, such as the authority to implement a test and, more importantly, to stop the administration of a test because of possible security concerns, must also be determined in advance. Decisions to temporarily stop using or retiring an online test can potentially delay recruiting initiatives or impact other groups, such as those responsible for proctored testing. The impact of such decisions must be carefully weighed by appropriately qualified individuals and standard protocols must be established. Additionally, authority and policies to grant re-tests or test exemptions due to technology issues and other such issues must also be identified.

12.2.4 Assessment Implementation

When deciding on the nature of the assessments to be implemented, the considerations identified in the previous three sections should be acknowledged prior to establishing which assessment an organisation may wish to administer online. Implementation experience and resource limitations may lead to the decision to use a simpler implementation of an assessment; for example, focussing on a test that is less prone to cheating and, thus, somewhat easier to implement. The development and validation of online assessments requires significant resources; therefore, when considering options for which assessment to implement online, the scope should extend beyond the existing assessments currently being used by the organisation, to include other variations of existing tests in use, as well as altogether new assessments of the psychological constructs that are to be measured. It is also important to carefully consider how the assessment will be used in the online context. For example, unproctored pre-selection tests provide substantial information, often with fewer implementation considerations. As a result, cheating may be less of a concern as online unproctored pre-selection test results may be verified later in the selection process with a proctored version of the test. Moreover, computer adaptive tests reduce the potential of item exposure and cheating, but they come at the cost of developing and validating a large item pool prior to use. When exploring assessment implementation options, consideration must be given to how test results from the online version will be validated.

12.2.4.1 The Special Case of Cognitive Ability Assessments

It is well recognised that cognitive ability is an important predictor of job performance [2], and consequently cognitive ability is a staple assessment in many selection processes, including the ones represented by this RTG. All nations responding to the RTG survey reported the use of one or more cognitive ability assessments in their applicant selection process (Chapter 8). The nature of most cognitive ability tests (i.e., the traditional fixed-item multiple choice response format with one correct answer) is also the reason why this test is one of the most challenging assessments to effectively implement in an online unproctored environment, as these tests are particularly vulnerable to compromises in test security, which influences the complexity of the implementation of an online assessment system. In other words, the nature of the test will determine the extent of the challenges associated with implementing an online assessment.

12.3 CONCLUDING REMARKS

For those exploring the implementation of an online personnel selection test, early success may be achieved through the implementation of a less complex test, and this experience may be leveraged to enable the implementation of a more complex test in the future.

12.4 REFERENCES

- [1] Fallaw, S.S., Kantrowitz, T.M., and Dawson, C.R. (2013). Global assessment trends report. SHL PreVisor.
- [2] Schmidt, F.L., and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2), 262-274.

Chapter 13 – DIVERSITY AND PERSONNEL SELECTION

LCol Soo Owens
Canadian Armed Forces
CANADA

Janne Hellerup Neilson
Danish Defence Ministry Personnel Agency
DENMARK

Wendy Darr and Maxime Grenier
Department of National Defence
CANADA

13.1 OVERVIEW

As alluded to in the introductory chapter of this report, diversity in the military workforce is becoming more vital not only in regard to recruiting personnel, but also in attracting those with the varied skills and competencies required for addressing complexity in work due to globalisation and technological advances. Consequently, personnel selection tools and processes must remain aligned with an organisation's diversity initiatives to facilitate the achievement of strategic diversity objectives. Unlike other NATO publications on diversity which examined topics related to the integration of certain groups into the workforce to improve diversity (e.g., TR-SAS-120), this chapter is focussed on issues pertinent to personnel selection. Following a brief examination of the concept of diversity, we explore the current state of military nations with respect to diversity-related personnel selection practices, highlight personnel selection considerations in the context of diversity, and provide insight into an emerging area of diversity that has implications for personnel selection.

13.2 THE CONCEPT OF DIVERSITY

Diversity within a workforce setting is a complex phenomenon, which is reflected in the multifold definitions of diversity. Overall, there seems to be two general approaches to defining workforce diversity within the literature. A narrow view defines workforce diversity only as differences based on the traditional categories such as race, colour, religion, gender, and national origin. A broader view argues that workforce diversity includes all the ways in which people can differ, which means that besides the traditional categories, it also includes differences in values, abilities, organisational function, work tenure, personality and so forth [11]. Others [14] have a more integrative perspective and suggest that diversity as an overall concept can be divided into two levels: surface-and deep-level. Surface-level diversity refers to readily visible characteristics (e.g., age, gender, race, ethnicity). Deep-level diversity, on the other hand, refers to non-observable characteristics such as working tenure, professional competence, education, values, and opinions. The latter is often overlooked when evaluating the effectiveness of recruitment or personnel selection practices, as in one discussion [26] of the underrepresentation of various racial groups and recruitment challenges in the United States' military.

While some research stresses the importance of considering both surface- and deep-levels of diversity [15], others argue that even simple categorical variables may not be an accurate reflection of surface-level indicators [25]. For example, the long-standing approach to measuring gender as a binary variable (i.e., man and woman) is now considered to be inadequate for capturing the spectrum along which one's gender identity could vary [48]. A systematic operationalisation of ethnicity has also been advocated, as there is no consensus on appropriate labels for diversity categories, whether they be ethnic groupings or biological sex differences [16]. Consequently, the measurement and operationalisation of diversity remains an issue, and it is beyond the scope of this chapter to cover or address. As much of the research on personnel selection focusses on traditional

categories of diversity variables (e.g., test score comparisons based on ethnic group categories [13]), our coverage of this topic is similarly limited to surface-level indicators of diversity, unless specified otherwise.

13.3 CURRENT STATE OF DIVERSITY IN MILITARY PERSONNEL SELECTION

To obtain insight into the extent to which diversity plays a role in military personnel selection practices, information was sought through a short survey of nation representatives participating in this RTG. Specifically, representatives were asked whether their military had a definition of diversity, and whether specific selection practices that reflect an attention to diversity (e.g., Research and Development [R&D] to monitor adverse impact, availability of multi-lingual assessments) exist within their organisation. The responses summarised herein are limited in that they reflect those of eight nations and should not be taken to represent the state of affairs across defence. Of those who responded, three representatives (Canada, Denmark, Germany) indicated that their military organisation had a definition of diversity at the time of the survey in 2018. With respect to specific selection practices that reflect attention to diversity, four representatives (Belgium, Canada, Finland, Germany) described practices such as evaluating for test score differences across gender, protected groups, and language, making selection tools (e.g., interviews) available in the nation's official languages, culture-fair test development practices, and developing procedures to address the needs of specific groups (e.g., those with disabilities). Overall, there appears to be variation in the levels of interest in diversity issues across nations.

13.4 IMPLICATIONS OF DIVERSITY FOR PERSONNEL SELECTION

Various research literatures have highlighted the unintended consequences that diversity can have on personnel selection practices. Studies on cognitive biases and decision-making show that subconscious heuristics and biases among recruiters play a considerable role in the processes that lead to workforce homogeneity or diversity [8], [22], [38]. One study [22] identified that one of the most robust predictors of a recruiter's assessment of a candidate during an interview is race similarity between the recruiter and the candidate. There is also evidence showing that migrants who express their unique cultural heritage are generally viewed by recruiters as having a lesser person-job match than migrants who associate more with the majority culture [17], [28], [38]. On the other hand, positive attitudes towards diversity among recruiters has been shown to positively affect the assessment and rating of a candidate belonging to a minority group [27].

Another body of research pertains to the diversity-validity dilemma, a label used to describe the pervasive finding of ethnic sub-group differences in some cognitive measures used in personnel selection [16], [38], [41]. Cognitive ability tests tend to be popular within the field of selection because they have been found to be the best single predictor of job outcomes [47]. However, evidence from several studies suggest that cognitive ability tests can result in adverse impact against minority groups, having a negative influence on minority group representation in the organisation [37], [43], [44]. Consequently, the dilemma implies that robust personnel selection efforts and diversity goals are not always aligned.

13.5 FAIRNESS CONSIDERATIONS IN PERSONNEL SELECTION

Although fairness to all applicant groups has been foundational to the development of defensible personnel selection practices, fairness considerations take on added significance as the applicant pool becomes more diverse. This becomes particularly challenging when applicants represent an intersection of diversity indicators such as age, ethnicity, and language (e.g., older woman from a minority ethnic group who is proficient only in a

non-native language). In addition, applicants who vary on non-demographic indicators related to academic achievement (e.g., learning disabilities) and information processing (e.g., attention deficit disorders) must also be accounted for when examining the fairness of personnel selection practices. The Standards for Educational and Psychological Testing (referred to as the *Standards* henceforth [2]) and the International Test Commission guidelines [29] outline a set of principles dedicated to ensuring fairness in assessments. Given the thorough coverage of this topic in these two sources, the goal of this section is to simply raise an awareness of the kinds of considerations that must be given in personnel selection practices to fully embrace applicant diversity.

Improving fairness in personnel selection implies improving accessibility for all diversity groups, where accessibility is described as ensuring that all applicants have an equal opportunity to demonstrate their level of skill or competence on a particular construct assessed by a test or assessment procedure [2]. In other words, efforts must be made to identify and remove variance that is unrelated to the construct, or construct-irrelevant variance, which may result from a particular diversity indicator. For example, a cognitive ability test containing questions with North American cultural content may be inaccessible or may introduce construct-irrelevant variance for applicants who have lived in a particular European country for most of their lives. Consequently, fairness in personnel selection practice requires applicant diversity characteristics to be considered at every stage of the selection R&D process: test development, test administration, and test score interpretation. Below, we highlight some considerations that must be given within each of these stages to ensure accessibility, particularly for diverse applicant groups.

13.5.1 Test Development

In their description of fairness in the design of assessments, the *Standards* [2] make reference to the principles of Universal Design (UD), a concept that was originally introduced in the architectural domain to create barrier-free buildings and facilities. UD is now well-known within the educational field where its principles are applied to create learning environments and assessments to ensure that all students have the same opportunity to express mastery of a construct during standardised testing [50]. The principles of UD are based on seven elements [50] which, when applied to the test development process, will ensure that diverse applicant characteristics are deliberately taken into account, including defining the construct(s) to be assessed by the test, identifying test content, response formats, and testing procedures [2]. Table 13-1 lists the seven elements and provides a brief description of general R&D considerations as they pertain to diverse applicant groups. These are taken from Canada's recent review into the application of UD principles in the development of the Canadian Armed Forces personnel selection assessments (cf., Ref. [6]). For more detailed considerations, particularly with respect to the development of tests for use with cultural and linguistic groups, readers are referred to the International Test Commission's guidelines [29] for the development of large-scale assessments for such groups.

13.5.2 Test Administration

The *Standards* [2] and the International Test Commission [29] describe principles that focus on the comparable treatment of all applicant groups during the test administration process. This includes minimising the introduction of construct-irrelevant variance in test scores due to, for example, personal biases of test administrators against certain diverse applicant groups, differential familiarity of certain applicant groups with the use of computerised technology, or the use of test instructions in a language that is not the primary language of the applicant (unless language is the construct assessed by the test). While standardised test administration procedures typically ensure equal treatment of all applicant groups, sometimes provisions or changes to the standardised test administration process are required to improve accessibility for some applicant groups. These provisions are often referred to as test accommodations or modifications. While both represent changes to or adaptations of the standardised testing protocol, test accommodations do not interfere with the construct assessed

by the test, whereas test modifications tend to change the construct assessed by the test [2]. Both could involve changes to the test setting (e.g., small quiet room versus group administration), response format (e.g., marking answers on test booklets versus bubble sheet), timing (e.g., allowing more time than the standard allotted time), content presentation (e.g., larger font, oral versus written presentation), or use of aids (e.g., translator, calculator) during testing [2]). As these changes have implications for test score interpretation (discussed in the next section), they must be accompanied by documented R&D evidence supporting their use. Such R&D may include consultations with professional experts, qualitative studies with members from a diverse group, and/or experimental research.

Table 13-1: R&D Considerations Based on the Elements of Universal Design.

Universal Design Element	R&D Considerations
Inclusive Assessment Population	<p>Deliberately recruit and engage diverse groups of individuals (representative of the applicant pool) during test development and/or field testing.</p> <p>Undertake accessibility evaluations and fairness reviews at the time of new test development/design by employing standard accessibility evaluation tools (e.g., Test Accessibility and Modification Inventory [TAMI], [5]) or using fairness review guidelines (e.g., Ref. [51]) to ensure the development of accessible test content/items.</p>
Precisely Defined Concepts	<p>Ensure that the operational definition of the primary construct of interest is clear, and that potential construct-irrelevant cognitive, sensory, emotional, social, and physical barriers (as they pertain to the applicant pool) can be identified and eliminated prior to test development.</p>
Accessible, Non-Biased Items	<p>Incorporate accessibility evaluation tools (e.g., TAMI, [5]) in the review of new assessments.</p> <p>Review current assessments (especially those that were developed a long time ago) to ensure accessibility for diverse applicant groups.</p> <p>When field testing of all assessment tools, deliberately include diverse groups representative of the applicant pool.</p> <p>For computer-based assessments, employ the use of existing frameworks (e.g., Accessible Portable Item Protocol [APIP], [49]) to ensure the transfer of accessible assessment content onto the delivery system.</p>
Amenable to Accommodations	<p>Should accessibility evaluations identify the need for test accommodations for some groups, determine the feasibility of such modifications (from a practical and legal standpoint), and conduct R&D to include such groups at the time of field testing, to ensure comparable measurement.</p> <p>Develop alternate (i.e., accommodated) assessment versions, if required, at the same time as standard assessment development.</p>

Universal Design Element	R&D Considerations
Simple, Clear, and Intuitive Instructions and Procedures	<p>Gather feedback, from diverse end-users, on test instructions and procedures during field testing to ensure that these are designed to be simple, clear, and presented in understandable language without excess wordiness.</p> <p>Consider making sample or practice items of all operational assessments available to applicants prior to completing operational assessments. This will facilitate understanding and comfort with the test administration procedures and format.</p>
Maximum Readability and Comprehensibility	<p>Employ accessibility tools and standards (e.g., TAMI, APIP) in the review of all assessment content.</p> <p>Ensure grade-level equivalence of all assessment content. According to test development guidelines, testing instructions/content should be at least two grade levels below the specified entry requirement [4].</p>
Maximum Legibility	<p>Examine the feasibility of employing established standards (e.g., APIP, World Wide Web Consortium; W3C, [51] accessibility standards) in the coding of computer-based items to ensure that assessment items and tests can be delivered in a standard format, with the potential to address specific test-taker needs.</p>

Note: Original source [6]

13.5.3 Test Score Interpretation

Ensuring fairness to all applicant groups requires evidence that supports the validity of test scores for applicant groups that vary in terms of language, gender, culture, or other attributes. As test scores in personnel selection are used to make inferences about applicants' level or standing on a construct, as well as their future performance on the job, evidence must be compiled to support both inferences [2]. The former is often supported by examinations showing invariance of constructs across various groups. These may include comparisons of the internal factor structure of the test, associations with other similar tests, and/or differential item functioning across various applicant groups [29]. Evidence for test score inferences about applicants' potential to perform on the job, on the other hand, involves examining the criterion-related validity of the selection assessment in each of the diverse groups, with evidence showing comparable validity across groups [2].

When comparable validity across diverse groups cannot be supported, efforts must be made to identify the underlying sources of these differences, to ensure that they result from construct-relevant rather than construct-irrelevant (e.g., applicant characteristics) factors [29]. When construct-irrelevant factors cannot be removed from an assessment, the assessment is typically adapted to a particular applicant group through test accommodations and/or modifications. However, accommodated or modified tests must be similarly supported by evidence to show that comparable inferences can be made to the original test about test scores from adapted assessments [2]. When an assessment is found to be valid, but adverse outcomes (not linked to construct-irrelevant factors) are found for some groups, decisions about the continued use of such assessments are typically addressed via policy, or through combining the results with those from other assessments. For example, the generation of composites based on the addition of non-cognitive assessments (e.g., personality) to cognitive ones has been presented as a strategy for addressing the validity-diversity dilemma [38], [39].

13.6 NEURODIVERSITY: AN EMERGING CONSIDERATION

This chapter would not be complete without discussing the concept of neurodiversity, a term that is used to represent a category of individuals who are neurologically different from the neurotypical population [7]. Neurodiversity first emerged in the 2000s as a movement to address the marginalisation of those individuals with autism spectrum disorders, with the goal of promoting the inclusion of those individuals within society and the acceptance of autism as a whole [30], [31]. Since then, some have suggested that the concept evolve to include other conditions, such as Attention Deficit Hyperactivity Disorder (ADHD), Tourette's syndrome, and mood disorders [7], [24]. However, much of the literature on neurodiversity focusses on individuals on the autism spectrum.

13.6.1 Neurodiversity in Organisations

Over the last decade, organisations have begun to recognise the many talents of neurodiverse individuals, and examples of their employment and productivity in organisations have been on the rise. These are showcased in a special issue of the *Journal of Management and Organization* [32]. Examples include an internship program for neurodivergent individuals at a large financial institution, Deutsche Bank [42], as well as a work experience program to determine such individuals' aptitude for performing in information technology (IT) roles at DXC Technology [12]. Microsoft also has a program through which it has successfully hired neurodiverse individuals for various roles such as software engineer, data analyst, or lab engineer [35]. An international effort that originated in Denmark is the Specialisterne Foundation, a concept and trademark that promotes the benefits of integrating the unique competencies of people on the autism spectrum in the workforce (see <https://specialisterne.com>). Branches are said to be present in 12 countries (e.g., Australia, Canada, France, Ireland, Singapore, Spain, United States). The program provides support to neurodiverse individuals and workforces to ensure inclusive environments, where individuals work in IT areas such as software testing, programming, or data entry.

13.6.2 Neurodiversity in Defence and Related Sectors

13.6.2.1 Israeli Defence Forces (IDF)

When speaking of neurodiversity and the military, the IDF has been the primary and lead actor in the world through its recruitment of neurodiverse individuals for its geospatial intelligence unit, Unit 9900. This unit, formed in the early 2010s through the Ro'im Rachok (looking ahead) program, was initially developed to allow Israeli civilians on the autism spectrum to be eligible for military service [45]. Of note, military service in the IDF is mandatory. Those selected through the Ro'im Rachok program undergo training and apprenticeship for six months, and then proceed to serve for a mandatory period of 1 year, with an option to extend their terms of service [23]. Neurodivergent soldiers were initially employed in Unit 9900 to analyse geographical data such as maps and aerial surveillance photos, because of their supposed superior attention to detail which is said to constitute a major advantage [33]. To address the unit's reconnaissance and target acquisition functions [1], recruits have also been called upon to acquire capabilities relating to intelligence collection in cyberspace, as well as software and system maintenance [33]. While Ro'im Rachok and Unit 9900 were designed to serve as force multipliers for the IDF, they were also meant to be a job program with the aim of providing autistic recruits with skills that would make them competitive in the workforce after their terms of service [33].

13.6.2.2 Australian Defence Organisation (ADO) and DXC Technology

Under the Dandelion program, the ADO and DXC Technology partnered to develop a strategy for the recruitment and integration of people on the autism spectrum, aiming to capitalise on what they refer to as superior analytical skills and work drives [3]. Recruits were trained solely as cybersecurity analysts to work within dedicated pods or units in a civilian capacity. The ADO sees the initiative as a way to boost its capabilities, while also developing the capacities of the recruit (social, technical, etc.) so that they may be more successful in their future professional endeavours and lives in general [18]. Outside the strict borders of Defence, similar partnerships also seem to have been made with other Australian government organisations, such as the National Australia Bank [20].

13.6.2.3 The United Kingdom

Both the Royal Navy and Metropolitan Police have shown interest in supporting existing personnel within their ranks. Although no program exists for the specific recruitment of neurodiverse people, some military members of the United Kingdom have written about the potential of using such individuals in a data driven environment [24]. Further developments may follow on within the Royal Navy given that their second highest ranking officer, Vice Admiral Hines, publicly revealed that he was “on the spectrum” in a 2021 interview [9]. Vice Admiral Hines highlighted the potential of neurodiverse individuals to think differently, a key requirement for competing with adversaries, and expressed his wishes in making the Royal Navy recruitment process more adapted to the needs of neurodiverse individuals [36].

13.6.3 Implications of Neurodiversity in Personnel Selection

Personnel selection practices will need to employ innovative approaches to address the specific needs of neurodiverse applicants. As discussed earlier, applying the principles of UD will likely yield assessments that are fair and accessible for this group. This may entail the modification of existing selection tools, and/or the modification of constructs assessed by a tool. In such cases, comparability of measurement, validity, or test score inferences with the original assessment and/or across other groups will most likely not be applicable, but the resulting adaptation will yield a more accurate representation of the skills and abilities of neurodiverse individuals [2]. In addition, given that symptomatic manifestations of autism vary across individuals on the spectrum [32], it is possible that assessments will have to be tailored to the meet the needs of a specific individual. Consequently, traditional standardised assessment tools (e.g., structured interviews, standardised tests) are likely to be inaccessible to individuals in this group. For example, neurodiverse individuals are characterised as being socially inept, lacking the interpersonal and communication skills required to interview successfully [7], [40]. Alternate assessment approaches such as group socialising sessions with managers [40], formulating task-specific or knowledge-based questions and providing sufficient time to analyse [10], self-reported questionnaires containing the skills and interests of neurodiverse individuals [21], and work samples (e.g., robotic construction or programming [40]) that allow neurodiverse individuals to demonstrate their skills on tasks that closely approximate those in the work setting are other examples of assessments that may be suitable for this group [34]. As the goal with all these assessments is to maximise person-job fit, the work tasks to be performed by neurodiverse individuals on the job must be clearly identified to determine appropriate assessment or training opportunities that will allow them to demonstrate their Knowledge, Skills, and Abilities (KSAs).

The selection of neurodiverse individuals is by no means an easy task and must be accompanied by supports or accommodations in the workplace to facilitate their integration and productivity. For example, a four-stage approach was proposed to identify neurodiverse talent for the cybersecurity environment [46]. It began with:

- 1) The specification of cyber-related KSAs;
- 2) The development of a gaming exercise within which to instruct or assess the KSAs;
- 3) The deployment of selected neurodiverse individuals within the workplace (along with necessary accommodations, training, and education to support these individuals as well as their managers and co-workers); and
- 4) The evaluation of the program to assess the economic or productivity benefits to the employer, but also KSA improvements and benefits for neurodiverse individuals.

Implementing such a program in the defence context may be wrought with additional challenges. However, given reports that there are about 177 men and women currently serving in the Royal Air Force who have been diagnosed with autism [9], military personnel selection practitioners may soon be tasked with examining their current selection protocols to ensure that they remain fair and inclusive to the neurodiverse group.

13.7 SUMMARY

Drawing upon the *Standards* and the International Test Commission guidelines, this chapter draws attention to the many considerations that must be given to test development, test administration, and test score interpretation in order to ensure fairness to diverse applicant groups in personnel selection. The principles of UD, when applied at the outset (i.e., at the time of test development), are likely to yield assessments that remove construct-irrelevant variance and ensure that personnel selection practices remain accessible to all applicant groups. While these solutions pertain primarily to personnel selection R&D, equal attention must be given to issues concerning the attitudes and cognitive/decision-making biases of those entrusted with the operational administration of a personnel selection process. Training to increase personnel awareness of cognitive biases [19] and programs that inculcate positive attitudes towards diversity [27] are equally important in ensuring fairness in personnel selection to all diverse applicant groups.

13.8 REFERENCES

- [1] Ahronheim, A. (2020). Meet the unit behind the scenes of the IDF's precision warfare. The Jerusalem Post. <https://www.jpost.com/israel-news/meet-the-unit-behind-the-scenes-of-the-idfs-precision-warfare-633617>
- [2] American Educational Research Association, American Psychological Association, and National Council for Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association
- [3] Austin, R.D., Fieldhouse, M., Mohan, A., and Quinn, P. (2017). Why the Australian Defence Organization is recruiting cyber analysts on the autism spectrum. Harvard Business Review. <https://hbr.org/2017/12/why-the-australian-defence-organization-is-recruiting-cyber-analysts-on-the-autism-spectrum>
- [4] Baranowski, R.A. (2006). Item editing and editorial review. In S.M. Downing and T.M. Haladyna (Eds.), Handbook of test development, 349-358. Routledge (Taylor and Francis Group).
- [5] Beddow, P.A., Kettler, R.J., and Elliott, S.N. (2008). Test accessibility and modification inventory. <http://peabody.vanderbilt.edu/tami.xml>

- [6] Boyes, F.A.J., Sobhani, S., and Darr, W. (2021). The applicability of universal design in Canadian Armed Forces selection (Director General Military Personnel Research and Analysis Scientific Report DRDC-RDDC-2021-R107). Ottawa, ON: Defence Research and Development Canada.
- [7] Brinzea, W. (2019). Encouraging neurodiversity in the evolving workforce – The next frontier to a diverse workplace. *Scientific Bulletin – Economic Sciences (Special Issue)*, 18, 13-25.
- [8] Brooks, M.E., Guidroz, A.M., and Chakrabarti, M. (2009). Distinction bias in applicant reactions to using diversity information in selection. *International Journal of Selection and Assessment*, 17, 377-390.
- [9] Brown, L. (2021). Vice Admiral Nick Hine: “To be blunt, autism made me a better naval officer.” Retrieved 4 Oct 2021 from <https://www.thetimes.co.uk/article/vice-admiral-nick-hine-to-be-blunt-autism-made-me-a-better-naval-officer-rvg8p8rxl>
- [10] Burnett, K. (2019). Embracing neurodiversity in the workplace. *Training Journal*. Retrieved 5 Oct 2021 from <https://www.trainingjournal.com/articles/features/tj-magazine-embracing-neurodiversity-workplace>
- [11] Carrell, M.R., Mann, E.E., and Sigler, T.H. (2006). Defining workforce diversity programs and practices in organizations: A longitudinal study. *Labor Law Journal*, 57(1), 5-12.
- [12] Carrero, J., Krzeminska, A., and Härtel, C.E.J. (2018). The DXC technology work experience program: disability-inclusive recruitment and selection in action. *Journal of Management & Organization* 25(4), 535-542. doi.org/10.1017/ jmo.2019.23
- [13] Cascio, W.F., and Outzz, J. (1995). Statistical implications of six methods of test score use in personnel selection. *Human Performance*, 8(3), 133-164. doi:10.1207/s15327043hup0803_2
- [14] Casper, W.J., Wayne, J.H., and Manegold, G.M. (2013). Who will we recruit? Targeting deep- and surface-level diversity with human resource policy advertising. *Human Resource Management* 52(3), 311-332. https://www.researchgate.net/journal/1099-050X_Human_Resource_Management
- [15] Cho, S., Kim, A., and Barak, M.E.M. (2017). Does diversity matter? Exploring workforce diversity, diversity management, and organizational performance in social enterprises. *Asian Social Work and Policy Review*, 11, 193-204.
- [16] De Soete, B., Lievens, F., and Druart, C. (2012). An update on the diversity-validity dilemma in personnel selection: A review. *Psychological Topics*, 21 (3), 399-424.
- [17] De Vroome, T., and Van Tubergen, F. (2010). The employment experience of refugees in the Netherlands. *International Migration Review*, 44, 376-403.
- [18] Department of Defence. (2016). Defence builds cyber capability with talents of people with autism. <https://www.defence.gov.au/annualreports/16-17/Features/CyberCapabilityTalentAutism.asp>
- [19] Derous, E., Nguyen, H-H., and Ryan, A.M. (2020). Reducing ethnic discrimination in resume-screening: A test of two training interventions. *European Journal of Work and Organizational Psychology*, 30(2), 225-239.

- [20] DXC Technology. (2021). Neurodiversity program develops unique skills and talents for National Australia Bank. Retrieved 5 Oct, 2021 from <https://dxc.com/au/en/insights/customer-stories/neuro-diversity-program-develops-unique-skills-and-talents-for-national-australia-bank>
- [21] Gal, E., Meir, A.B., and Katz, N. (2013). Brief report: Development and reliability of the Autism Work Skills Questionnaire (AWSQ). *American Journal of Occupational Therapy*, 67, e1-e5.
- [22] Goldberg, C. (2005). Relational demography and similarity-attraction in interview assessments and subsequent offer decisions: Are we missing something? *Group & Organization Management*, 30, 597-624.
- [23] Goldenberg, T. (2016). For autistic soldiers, army service opens door to integration. Retrieved 5 Oct 2021 from <https://www.timesofisrael.com/for-autistic-soldiers-army-service-opens-door-to-integration/>
- [24] Gordon, J. (2020). Neurodiversity in defence and a chaotic world. Wavell Room. <https://wavellroom.com/2020/12/30/neurodiversity-defence-chaos-advantage-opportunity/>
- [25] Hall, Jr., E.H., and St. John., C.H. (1994). A methodological note on the measurement of diversity. *Strategic Management Journal*, 15, 153-168.
- [26] Harris, G.L.A. (2009) Recruiting, retention, and race in the military. *International Journal of Public Administration*, 32 (10), 803-828, DOI: 10.1080/01900690902905485
- [27] Hofhuis, J., Van der Zee, K., and Otten, S. (2016). Dealing with differences: The impact of perceived diversity outcomes on selection and assessment of minority candidates. *The International Journal of Human Resource Management*, 27(12), 1319-1339.
- [28] Horverak, J.G., Bye, H.H., Sandal, G.M., and Pallesen, S. (2013). Managers' evaluations of immigrant job applicants: the influence of acculturation strategy on perceived person- organization fit (p-o fit) and hiring outcome. *Journal of Cross-Cultural Psychology*, 44, 46-60.
- [29] International Test Commission. (2018). ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations. www.InTestCom.org
- [30] Kapp, S.K. (Ed.). (2020). Autistic community and the neurodiversity movement: Stories from the frontline. doi.org/10.1007/978-981-13-8437-0.
- [31] Kapp, S.K., Gillespie-Lynch, K., Sherman, L.E., and Hutman, T. (2013). Deficit, difference, or both? Autism and neurodiversity. *Developmental Psychology*, 49(1), 59-71. doi.org/10.1037/a0028353 17
- [32] Krzeminska, A., Austin, R.D., Bruyere, S.M., and Hedley, D. (2019). The advantages and challenges of neurodiversity employment in organizations. *Journal of Management and Organization*, 25(4), 453-463. doi:10.1017/jmo.2019.58
- [33] Kushner, D. (2019). Serving on the spectrum: The Israeli Army's Roim Rachok program is bigger than the military. *Esquire*. <https://www.esquire.com/news-politics/a26454556/roim-rachok-israeli-army-autism-program/>

- [34] Marcy, R.T., and Bayati, A. (2020). How I-O psychology can help in the selection and development of neurodiverse employees. Society for Industrial and Organizational Psychology White Paper Series. <https://www.siop.org/Portals/84/docs/White%20Papers/neurodiverse.pdf>
- [35] Microsoft. (2021). Microsoft neurodiversity hiring program and FAQ. Retrieved 4 Oct 2021 from <https://microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring>
- [36] National Autistic Society. (2021). Stories from the spectrum: Vice Admiral Nick Hine. National Autistic Society. Retrieved 5 Oct 2021 from <https://www.autism.org.uk/advice-and-guidance/stories/stories-from-the-spectrum-nick-hine>
- [37] Ng, E.S., and Sears, G.J. (2010). The effect of adverse impact in selection practices on organizational diversity: A field study. *The International Journal of Human Resource Management*, 21, 1454-1471.
- [38] Ployhart, R.E., and Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology*, 61, 153-172.
- [39] Potosky, D., Bobko, P., and Roth, P.L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment*, 13, 304-315.
- [40] Pricope, M., Mazilu, S., and Popa, F. (2020, April 23 – 24). Online tools for introducing neurodiversity in job interviews with engineering students. International Scientific Conference eLearning and Software for Education, Bucharest. doi:10.12753/2066-026X-20-233
- [41] Pyburn, Jr., K.M., Ployhart, R.E., and Kravitz, D. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61(1), 143-151. <https://doi.org/10.1111/j.1744-6570.2008.00108.x>
- [42] Remington, A., and Pellicano, E. (2018). ‘Sometimes you just need someone to take a chance on you’: An internship programme for autistic graduates at Deutsche Bank, UK. *Journal of Management and Organization* 25(4), 516-534, doi:10.1017/ jmo.2018.66
- [43] Roth, P.L., Bevier, C.A., Bobko, P., Switzer III, F.S., and Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, (2), 297-330.
- [44] Roth, P.L., Huffcutt, A.I., and Bobko, P. (2003). Ethnic group differences in measures of job performance: a new meta-analysis. *Journal of Applied Psychology*, 88(4), 694-706.
- [45] Sales, B. (2015). Soldiers with autism take on key roles in IDF. *The Times of Israel*. <http://www.timesofisrael.com/soldiers-with-autism-take-on-key-roles-in-idf/>
- [46] Scanlan, J., Watters, P.A., Eddy, A., Fieldhouse, M., Thomas, T., Fung, L., Tan, T., Girdler, S., and Chen, Y. P. (2020). Neurodiverse knowledge, skills, and ability assessment for cyber security. Australasian Conference on Information Systems, Wellington, New Zealand. <https://researchgate.net/publication/350964865>

- [47] Schmidt, F.L., and Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings, *Psychological Bulletin*, 124, 262-274.
- [48] Statistics Canada. (2020). Sex and gender. Retrieved 22 Sept 2021 from <https://www12.statcan.gc.ca/census-recensement/2021/road2021-chemin2021/fs-fi/sex-and-gender.cfm>
- [49] Stone, E., Laitusis, C.C., and Cook, L.L. (2016). Increasing the accessibility of assessments through technology. In F. Drasgow (Ed.), *Technology and Testing: Improving Educational and Psychological Measurement*, 217-234. Routledge Taylor & Francis Group.
- [50] Thompson, S., Thurlow, M., and Malouf, D.B. (2004). Creating better tests for everyone through universally designed assessments. *Journal of Applied Testing Technology*, 6(1), 1-15.
- [51] World Wide Web Consortium. (2008). Web content accessibility guidelines (WCAG) 2.0. <https://www.w3.org/TR/WCAG20/#keyboard-operation>
- [52] Zieky, M. (2006). Fairness review in assessment. In S.M. Downing, S.M., and T.M. Haladyna (Eds.). *Handbook of test development*. Routledge (Taylor and Francis Group).

Chapter 14 – ADVANCES IN MILITARY PERSONNEL SELECTION: NEXT STEPS

The information contained in this report represents the culmination of a collaborative effort to understand the state of affairs with respect to three divergent topics that are of significance to military personnel selection across nations. Nation representatives have had the opportunity to learn about each other's respective military personnel selection systems, current efforts, and future plans. In all three areas of focus (i.e., integrity, online testing, and diversity), nations appear to be at different phases in their efforts to advance their respective military selection procedures and processes. This report provides insight into some of these efforts. In the area of online testing, for example, commonalities in online implementation practices, as well as challenges that emerge from these implementations, are evident across nations. In the area of integrity testing, progress appears to be limited by a lack of consistency in understanding integrity, which has impacted its consistent assessment across nations. Diversity remains an area of significance to many military nations, but the scope of initiatives to address diversity-related issues in military personnel selection varies across nations.

Moving forward, as nations embark on initiatives to address requirements specific to their own organisation, there is an opportunity to learn from the experiences gained by other nations. As explained in Chapter 1, the ever-changing environment will require each nation to adapt and innovate their selection procedures in order to remain competitive. The assessment of applicants to our militaries is changing and becoming more complex. The consequences of doing it wrong are considerable and can range from the placement of individuals into less suitable occupations to legal challenges. The work of this RTG cannot cease with the publication of this report. There are several sub-areas of interest that will benefit from collaborative sharing of information. At the present time, some of these areas are listed below, but may include new (currently unknown) areas as well.

Online Testing

- Online software implementation
- Security
- Assessment management policy
- Assessment implementation
- Advanced online testing software applications (e.g., machine learning and artificial intelligence).

Integrity Testing

- Operationalisation or assessment of integrity as defined in Chapter 5
- Emerging approaches to assessing integrity

Diversity

- Application of universal design principles (see Chapter 13) in the development of selection tools

In closing, we reiterate that in order to sustain advancements in military personnel selection, there is a need to leverage the collective knowledge, experience, and expertise of nations on a continual basis. We, therefore, invite and encourage future collaborations among nations, but also call for the establishment of planned technical

activities within NATO STO in the form of events (e.g., symposia, workshops, specialist meetings); educational activities (e.g., instructor-led lecture series) to facilitate the exchange of information, focussed discussions, and state-of-the art knowledge; or exploratory teams around a particular focus area (e.g., emerging technologies).

Annex A – ONLINE TESTING: STANDARDISED TERMINOLOGY

When the RTG first met and started discussing various aspects of military selection testing, it quickly became clear that common terms were used differently by the representatives of the different nations. To improve communication, for the purposes of this report, the following definitions have been agreed upon:

- **Computer-based testing or digital testing:** This refers to any test that is delivered by a computer.
- **Internet-based testing or online testing:** Any test administered through the internet.
- **Mobile testing:** Completion of an internet-based test on a cellular phone or tablet.
- **Pre-selection test:** A screening test, often used early in a job applicant selection process, with the goal of filtering out candidates who do not match the job profile.
- **Personnel selection test:** A systematic and standardized assessment procedure for measuring job applicants' abilities, behaviours, or other qualities that is used in a selection process.
- **Proctored testing:** A testing environment where someone is either supervising or monitoring a test taker as they complete a test, either directly or via a webcam to reduce the possibility of cheating or recording test items.
- **Unproctored Internet Testing (UIT):** A situation where candidates are not monitored as they complete a test.



Annex B – ONLINE TESTING BENCHMARKING SURVEY (2018)

- Response ID.
- Which nation do you represent?
- Which nation do you represent? [Other].
- What is your name?
- What is your email address?
- What is the name of the test?
- What category does this test fall into?
- What category does this test fall into? [Other].
- What is this test?
- Please indicate when this test was first implemented.
- Please indicate when you plan on implementing this test.
- Briefly describe what the test is designed to evaluate and the nature of the test items. (e.g., cognitive ability test with traditional problem-solving test items. In total there are 50 test items. Test takers choose one of 5 multiple choice options. Test items are scored as one point for each correct answer.)
- Please provide information about how the content for your test is developed. For example, is it developed “in house”, obtained through private test vendors, or a combination? Please provide details.
- How is the test administered? [Digital test (not online) – proctored].
- How is the test administered? [Digital test (not online) – unproctored].
- How is the test administered? [Online test – proctored].
- How is the test administered? [Online test – unproctored].
- How is the test administered? [Paper-based test – proctored (as a backup in case of computer failure)].
- How is the test administered? [Paper-based test – unproctored (as a backup in case of computer failure)].
- If the test is administered in more than one way (more than one box was checked in the previous question), please identify approximately what proportion of the applicants complete the test using each method.
- Can this test be completed using a: [PC / laptop computer]?
- Can this test be completed using a: [Smartphone]?
- Can this test be completed using a: [Tablet]?
- Can this test be completed using a: [Other]?

ANNEX B – ONLINE TESTING BENCHMARKING SURVEY (2018)

- Please provide information on how the software and digital content (multimedia files, etc.) are developed for your test. Is the test developed “in house”, obtained through private test vendors, or a combination? Please provide details.
- Briefly describe the test administration process. Include information on how candidates are invited to take tests, how the test taker’s identity is authenticated, at what stage in the recruiting process does the candidate complete the test, and is feedback given to the candidate on their performance? If the process is the same as for a previous testing process you have described, simply state this, and make sure to identify the previous testing process.
- Depending upon the type of test there could be concerns over candidates cheating and test content security. Please identify such concerns and how they are addressed. If the concerns are the same as for a previous test you have described, simply state this, and make sure to identify the previous test.
- Aside from issues related to candidates cheating and test security, from technical perspective, how satisfied is your selection centre with the current test administration process? (consider issues such as system reliability, ease of making changes to software, and so on).
- Please indicate why you are satisfied or unsatisfied with the test administration process.
- Please identify any aspects of the test administration process you are considering changing.
- Is there a proctored follow-up to validate the unproctored test results?
- How is the follow-up test implemented?
- Are you currently using or planning to implement additional tests?

Annex C – ONLINE TESTING BENCHMARKING SURVEY (2021 UPDATE)

Nation:

Completed by:

Email address:

As it has been a while since the original survey, and as the COVID situation might have affected both your selection procedures and the development and/or use of online tests, please respond to the questions below.

Q1. Has the COVID situation affected any of your selection processes?

- Yes [] No []
If yes, please describe below:
- Selection process (which selection process):
- Challenges:
- Solutions:
- Experience with solution:
- Selection tool (which selection process):
- Challenges:
- Solutions:
- Experience with solution:

Q2. Has any recent development (e.g., COVID, tech development, org changes, economy, etc.) affected your use/development of **online tests**?

- Yes [] No []
If yes, please describe below:
- How?

Q3. Have you, since the survey of 2018, completed any developments to your tests, or have new insights you can share regarding online tests?

- Yes [] No []
If yes,
- New developments:
- New insight:
- Other comments:
- Please share:



REPORT DOCUMENTATION PAGE			
1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	STO-TR-HFM-290 AC/323(HFM-290)TP/1076	ISBN 978-92-837-2394-3	PUBLIC RELEASE
5. Originator	Science and Technology Organization North Atlantic Treaty Organization BP 25, F-92201 Neuilly-sur-Seine Cedex, France		
6. Title	Advances in Military Personnel Selection		
7. Presented at/Sponsored by	Final report of Task Group HFM-290.		
8. Author(s)/Editor(s)	Multiple	9. Date	May 2022
10. Author's/Editor's Address	Multiple	11. Pages	182
12. Distribution Statement	There are no restrictions on the distribution of this document. Information about the availability of this and other STO unclassified publications is given on the back cover.		
13. Keywords/Descriptors	Computerized testing; Diversity; Integrity; Neurodiversity; Online testing; Personnel selection		
14. Abstract	<p>This report explores topics within the umbrella of personnel selection, to identify current trends, challenges, and developments in the field. The focus is on three areas that have significance in today's defence context: integrity, online/computerised testing, and diversity. With respect to integrity, this report provides insight into the history of integrity testing in the military; examines the current state of integrity testing practice and research across participating nations; summarizes commercial integrity tests and emerging approaches to assessing integrity; and provides a framework for defining integrity in the military for the purposes of its assessment in military personnel selection. On the topic of online/computerized testing, it examines the selection testing practices of participating nations, in addition to exploring psychosocial factors, testing security challenges, options for online test administration, and the benefits to personnel selection that can result from technology implementation. In the area of diversity, the concept of diversity, fairness and inclusivity considerations in personnel selection practices, and neurodiversity initiatives within defence are explored. Drawing upon the current state of practice and research, this report hopes to raise awareness of and provide guidance on topics that deserve consideration in military personnel selection.</p>		





BP 25

F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs.o.nato.int



**DIFFUSION DES PUBLICATIONS
STO NON CLASSIFIEES**

Les publications de l'AGARD, de la RTO et de la STO peuvent parfois être obtenues auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la STO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus soit à titre personnel, soit au nom de votre organisation, sur la liste d'envoi.

Les publications de la STO, de la RTO et de l'AGARD sont également en vente auprès des agences de vente indiquées ci-dessous.

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivi du numéro de série. Des informations analogues, telles que le titre et la date de publication sont souhaitables.

Si vous souhaitez recevoir une notification électronique de la disponibilité des rapports de la STO au fur et à mesure de leur publication, vous pouvez consulter notre site Web (<http://www.sto.nato.int/>) et vous abonner à ce service.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE

Streitkräfteamt / Abteilung III
Fachinformationszentrum der Bundeswehr (FIZBw)
Gorch-Fock-Straße 7, D-53229 Bonn

BELGIQUE

Royal High Institute for Defence – KHID/IRSD/RHID
Management of Scientific & Technological Research
for Defence, National STO Coordinator
Royal Military Academy – Campus Renaissance
Renaissancelaan 30, 1000 Bruxelles

BULGARIE

Ministry of Defence
Defence Institute "Prof. Tsvetan Lazarov"
"Tsvetan Lazarov" bul no.2
1592 Sofia

CANADA

DGSIST 2
Recherche et développement pour la défense Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

DANEMARK

Danish Acquisition and Logistics Organization
(DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESPAGNE

Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

ESTONIE

Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

ETATS-UNIS

Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72
92322 Châtillon Cedex

GRECE (Correspondant)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HONGRIE

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALIE

Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport "Comparto A"
Via di Centocelle, 301
00175, Rome

LUXEMBOURG

Voir Belgique

NORVEGE

Norwegian Defence Research
Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

POLOGNE

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
SDFA – Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE

Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

ROUMANIE

Romanian National Distribution
Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

ROYAUME-UNI

Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down
Salisbury SP4 0JQ

SLOVAQUIE

Akadémia ozbrojených síl gen.
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIE

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

TURQUIE

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi
Başkanlığı
06650 Bakanlıklar – Ankara

AGENCES DE VENTE

**The British Library Document
Supply Centre**
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
ROYAUME-UNI

**Canada Institute for Scientific and
Technical Information (CISTI)**
National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Les demandes de documents STO, RTO ou AGARD doivent comporter la dénomination « STO », « RTO » ou « AGARD » selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications STO, RTO et AGARD figurent dans le « NTIS Publications Database » (<http://www.ntis.gov>).



BP 25
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@cs.o.nato.int



**DISTRIBUTION OF UNCLASSIFIED
STO PUBLICATIONS**

AGARD, RTO & STO publications are sometimes available from the National Distribution Centres listed below. If you wish to receive all STO reports, or just those relating to one or more specific STO Panels, they may be willing to include you (or your Organisation) in their distribution.

STO, RTO and AGARD reports may also be purchased from the Sales Agencies listed below.

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number. Collateral information such as title and publication date is desirable.

If you wish to receive electronic notification of STO reports as they are published, please visit our website (<http://www.sto.nato.int/>) from where you can register for this service.

NATIONAL DISTRIBUTION CENTRES

BELGIUM

Royal High Institute for Defence –
KHID/IRSD/RHID
Management of Scientific & Technological
Research for Defence, National STO
Coordinator
Royal Military Academy – Campus
Renaissance
Renaissancelaan 30
1000 Brussels

BULGARIA

Ministry of Defence
Defence Institute “Prof. Tsvetan Lazarov”
“Tsvetan Lazarov” bul no.2
1592 Sofia

CANADA

DSTKIM 2
Defence Research and Development Canada
60 Moodie Drive (7N-1-F20)
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC

Vojenský technický ústav s.p.
CZ Distribution Information Centre
Mladoboleslavská 944
PO Box 18
197 06 Praha 9

DENMARK

Danish Acquisition and Logistics Organization
(DALO)
Lautrupbjerg 1-5
2750 Ballerup

ESTONIA

Estonian National Defence College
Centre for Applied Research
Riia str 12
Tartu 51013

FRANCE

O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc – BP 72
92322 Châtillon Cedex

GERMANY

Streitkräfteamt / Abteilung III
Fachinformationszentrum der
Bundeswehr (FIZBw)
Gorch-Fock-Straße 7
D-53229 Bonn

GREECE (Point of Contact)

Defence Industry & Research General
Directorate, Research Directorate
Fakinos Base Camp, S.T.G. 1020
Holargos, Athens

HUNGARY

Hungarian Ministry of Defence
Development and Logistics Agency
P.O.B. 25
H-1885 Budapest

ITALY

Ten Col Renato NARO
Capo servizio Gestione della Conoscenza
F. Baracca Military Airport “Comparto A”
Via di Centocelle, 301
00175, Rome

LUXEMBOURG

See Belgium

NETHERLANDS

Royal Netherlands Military
Academy Library
P.O. Box 90.002
4800 PA Breda

NORWAY

Norwegian Defence Research
Establishment, Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND

Centralna Biblioteka Wojskowa
ul. Ostrobramska 109
04-041 Warszawa

PORTUGAL

Estado Maior da Força Aérea
S DFA – Centro de Documentação
Alfragide
P-2720 Amadora

ROMANIA

Romanian National Distribution Centre
Armaments Department
9-11, Drumul Taberei Street
Sector 6
061353 Bucharest

SLOVAKIA

Akadémia ozbrojených síl gen
M.R. Štefánika, Distribučné a
informačné stredisko STO
Demänová 393
031 01 Liptovský Mikuláš 1

SLOVENIA

Ministry of Defence
Central Registry for EU & NATO
Vojkova 55
1000 Ljubljana

SPAIN

Área de Cooperación Internacional en I+D
SDGPLATIN (DGAM)
C/ Arturo Soria 289
28033 Madrid

TURKEY

Milli Savunma Bakanlığı (MSB)
ARGE ve Teknoloji Dairesi Başkanlığı
06650 Bakanlıklar – Ankara

UNITED KINGDOM

Dstl Records Centre
Rm G02, ISAT F, Building 5
Dstl Porton Down, Salisbury SP4 0JQ

UNITED STATES

Defense Technical Information Center
8725 John J. Kingman Road
Fort Belvoir, VA 22060-6218

SALES AGENCIES

The British Library Document Supply Centre

Boston Spa, Wetherby
West Yorkshire LS23 7BQ
UNITED KINGDOM

Canada Institute for Scientific and Technical Information (CISTI)

National Research Council Acquisitions
Montreal Road, Building M-55
Ottawa, Ontario K1A 0S2
CANADA

Requests for STO, RTO or AGARD documents should include the word 'STO', 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of STO, RTO and AGARD publications are given in “NTIS Publications Database” (<http://www.ntis.gov>).