

ORIGINAL ARTICLE

Test Usage and Quality Awareness among Norwegian-Psychologists: The EFPA Test Use Survey 2020

EFPA test survey 2020 in Norway

Cato Grønnerød¹, Mikael Julius Sømhovd² og Rudi Myrvang³¹Department of Psychology, University of Oslo, Norway²Norwegian Health Directorate, Oslo, Norway³Norwegian Psychological Association, Oslo, Norway

Published: 07.03.2025

The authors declare no conflicts of interest

CONTACT: cato.gronnerod@psykomatikk.no



Introduction: We report results from the Norwegian segment of the *European Federation for Psychologists' Associations'* (EFPA) 2020 decennial survey of psychologists' attitudes towards tests and test use. Previous surveys have shown that psychologists have positive attitudes towards testing and are concerned about the quality of tests and assessment practice.

Methods: We analysed the responses of 1,523 *Norwegian Psychologists' Association* members to a 32-item online questionnaire with a five-point Likert scale (20% response rate). We compared the results with the 2017 test use survey using descriptive analyses, and we performed a confirmatory factor analysis.

Results: The results show that the respondents were generally satisfied with student training and valued tests as a source of information; however, they were unsatisfied with information about test quality. The regulation of tests and test use was a major concern, both regarding general regulatory frameworks and restrictions on test use depending on test user qualifications. The survey indicated that cognitive measures were the most widespread, although the respondents listed a wide variety of tests they were using.

Conclusions: Our results confirm findings from earlier studies that psychologists are positive towards testing but have clear concerns about both quality and test regulation. They request more information regarding test quality and express that test use should be regulated based on test user competence.

Keywords: testing, test use, psychological assessment, personality assessment, clinical assessment



Concerns over test quality and test user competence in psychological practice have been raised both internationally (Evers, 2012; Evers et al., 2013; Muñiz et al., 2001) and in Norway (Ryder, 2021a, 2021b; Vaskinn et al., 2010; Vaskinn & Egeland, 2012). Inappropriate test use may result in inadequate or unsuitable treatment, loss of benefits or unfair legal sanctions, among other risks (Handler & Meyer, 1998). However, the question of how practising psychologists should ensure that the tests they utilise are of sufficient quality is a complex issue. Some researchers have called for a system to validate tests (Mihura et al., 2019), and in Norway, efforts have been made to establish such a system for tests targeting children and adolescents (Kornør, 2015). Test quality should be evaluated internationally rather than through independent national reviews of the same research—except for aspects that require local adaptation.

Furthermore, a certain level of test user competence must be mandated for the application of tests in clinical practice. The *International Test Commission* (ITC) is an international body that advocates for test quality and test user competence, working closely with national psychologists' associations, including the *Norwegian Psychologists' Association*. The ITC issues guidelines for various aspects of test use, and guidelines pertaining to test use and user competence are included in the *Norwegian Psychologists' Association's* ethical code of conduct (International Test Commission, 2001). Additional guidelines with more limited regional scopes are also available (American Educational Research Association et al., 2014; European Federation of Psychologists' Associations, 2023; Krishnamurthy et al., 2022). The question remains as to how these guidelines can be effectively implemented in clinical practice.

To work towards the implementation of these guidelines, we must also understand the attitudes of psychologists towards testing and test user qualifications, as well as how and to what extent they employ tests and inventories in their daily clinical work. The *European Federation for Psychologists' Associations* (EFPA) conducts a decennial survey of European psychologists' attitudes towards tests and test use (Evers et al., 2017; Muñiz et al., 2001). Recently, the ITC has also collaborated in this endeavour.

Test use has special significance because it involves more than clinical skills (Grønnerød, 2024); it also draws on knowledge about the properties of the instruments, including stringent standardisation, reliability, validity and context-relevant norms. If test quality properties are lacking, this can only be partially mitigated by a competent clinician. Research clearly shows that judgments based on actuarial data from tests are superior to those based on free clinical judgment (Grove et

al., 2000; Ægisdóttir et al., 2006). Psychologists should therefore be encouraged to maintain and update their practical and theoretical competencies and knowledge regarding test quality, test use and developments in psychometrics. Nevertheless, learning assessment skills is challenging and requires a combination of various psychological competencies and continued practice (Grønnerød, 2024; Handler & Meyer, 1998).



Scientific interest in assessment instruments and their quality has increased in recent years (Evers, 2012; Greiff, 2017). However, it is less apparent that scientific production and theoretical developments are being absorbed and integrated by practitioners into their attitudes and practices. The volume of scientific studies relevant to any subfield of professional psychology is substantial, and this is no less true for assessment practitioners. The Norwegian segment of the EFPA 2009 data reflected similar concerns. Vaskinn et al. (2010) concluded that frequent test use is connected to increased concern for test quality. Still, though Norwegian users expressed concern about test quality, they checked it less frequently than their concerns would suggest. The use of unofficial test copies was also problematic (Vaskinn & Egeland, 2012), as was the lack of proper translations and local norms.

The European psychologists responding to both the 2000 and 2009 surveys (Evers et al., 2017; Muñoz et al., 2001) indicated that the field requires more regulation from professional and/or government bodies. Both EFPA's Standing Committee on Tests and Testing (SCTT) and the ITC are mandated to promote competent testing practices and assess test quality. Survey results from 2000 and 2009 support these mandates and underpin professional testing policies, yet many respondents expressed concerns that both the tests and their use fall short of required standards.

Testing practices have changed over the last couple of decades with the rise of internet-based testing. As early as 2006, the ITC published guidelines on computer-based and internet-based testing (International Test Commission, 2006). However, it remains unclear as to how far has this field advanced since then. Our impression is that computer-based testing has not fully replaced traditional paper-and-pencil tests. We need to gain a better understanding of why this is the case and the extent to which it is due to psychologists' attitudes towards this 'new' technology.

Heading into a new decade, the EFPA organised the 2020 test survey to follow up on the previous survey and assess whether test usage and attitudes have changed. The *Norwegian Psychological Association* distributed the survey among its members in Norway. In this article, we present the results from the Norwegian data, aiming to examine the following:

- 1 How do psychologists evaluate their training and information access?
- 2 What are psychologists' general attitudes towards test quality and test use?
- 3 What are psychologists' attitudes to test regulation?
- 4 What are psychologists' attitudes towards internet-based testing?
- 5 Which tests are predominantly used?
- 6 Are the a priori subscales found by Evers (2012) replicated in our data?



This last question could aid in interpreting broader groups of concerns and may guide future revisions of the questionnaire.

Methods

Procedure

We distributed the survey by email to 7,282 members of the *Norwegian Psychological Association* in mid-January 2020. The list included student members. The survey consisted of 32 questions (Table 1) regarding test use and attitudes towards tests and testing, which were answered on a five-point Likert scale, ranging from 1 (Totally disagree) to 5 (Totally agree). The survey text was translated from English to Norwegian and quality controlled by the *Norwegian Psychological Association's Board of Test Policy*. The survey was administered through an online platform. In addition to questions targeting their attitudes, respondents were asked to list their three most frequently used tests. No personal information was collected aside from age, gender and the professional field in which the individual was currently employed.

Respondents

When the survey had closed, 1,831 members (approximately 24%) had responded (1,269 women [69.6%], 553 men [30.2%], 5 with other gender identities, and 4 missing). Of these, 1,523 completed the questionnaires, resulting in an overall completion rate of approximately 20%. Most of the attrition occurred after the demographic information page; Table 1 shows the variations in responses for subsequent questions.

Analyses



First, we analyse the background differences between completers and non-completers. We also report descriptive statistics, categorising the respondents by gender, age and professional work domain area. We then examine the differences in demographic variables between test groups using *t*-tests and chi-square tests as appropriate for the respective variables. Variables are primarily presented and analysed as sum scores. Figures (which can be found at https://osf.io/dh7xs/?view_only=0b3b42a93b1244999b1374f99965e6e6) show that the distribution is approximately symmetric and the means are moderate. We ran correlational analyses for age and a one-way ANOVA for work area, with all correlations significant at $p < .001$. Our sum score analyses adhered to Evers et al.'s (2017) presentation of the 2009 data, which enabled us to conduct a supplementary confirmatory factor analysis. In this analysis, we transformed the questionnaire into five a priori subscales (*Concern*, *Regulations*, *Internet*, *Appreciation* and *Knowledge*) – and analysed their properties, as identified by Evers et al. The subscales are defined as follows: *Concern* – the degree to which psychologists are concerned about problems related to test use; *Regulations* – the extent of their support for stricter regulations on tests and testing; *Internet* – their positivity towards the value and effectiveness of internet-based testing; *Appreciation* – their belief that tests are valuable tools for practitioners; and *Knowledge* – their endorsement of training as a prerequisite for appropriate test use.

Results

Preliminary Analyses

Our data suggest no statistical differences in gender between completers and non-completers ($\chi^2 = .52, p = .470$). Due to non-normality ($W = .96, p < .001$), we utilised the *Mann–Whitney U test* which revealed a significant ranked age difference ($U = 2.38, p = .020$) between completers ($Mdn = 42$) and non-completers ($Mdn = 39$). Finally, professional domain differed between completers and non-completers ($\chi^2_{(3)} = 28.28, p < .001$). Adjusted residual for the *Clinical/Health* group (-4.7) and the *Other* group (5.5) indicated more completers in the former and fewer completers in the latter. The adjusted residual Z-values for both the *Work* and *Education* groups were below 1.96, suggesting a non-significant contribution to the χ^2 estimate. Correlations with age displayed effect sizes ranging from small to medium ($r = -.22$ to $.23$).



The mean age of respondents was 43.6 years ($SD = 12.76$, range 23 – 85, $Mdn = 41$). Male respondents were significantly older than female respondents ($M = 47.0$, $SD = 13.7$ versus $M = 42.2$, $SD = 12.1$; $p < .000$). Most respondents worked in specialist health care for adults (29.3%) and children (16.5%) or other/several areas (12.7%). Private practice represented 9.4% of the sample, and 6.1% worked as neuropsychologists. Women were overrepresented in all areas except work psychology, reflecting the overall gender composition of practising Norwegian psychologists. Table 1 shows the mean response level, standard deviation and skewness for each of the 32 individual questions analysed in the following.

Table 1

Questionnaire Responses, Mean, Standard Deviation and Skew, Grouped by Topic

No.	Question	N	Mean	SD	Skew
Information and Training					
1	The training received in psychology bachelor's degree courses is sufficient for the correct use of most tests	1,520	2.25	0.989	0.497
2	The training received in psychology master's degree courses is sufficient for the correct use of most tests	1,515	3.08	0.996	-0.162
4	Professionals are provided with sufficient information (independent reviews, research, documentation etc.) on the quality of tests published in my country	1,515	2.43	0.935	0.301
6	My current knowledge with regards to tests is basically that which I learned in my psychology degree course	1,516	2.47	1.244	0.537
Regulation					
3	The European Federation of Psychologists Associations (EFPA) should establish a European system to accredit the certification of test users	1,511	3.20	1.108	-0.157
8	The use of psychological tests should be restricted to qualified psychologists	1,510	3.89	1.063	-0.802
9	While non-psychologists may administer and score tests, interpretation and feedback should be restricted to psychologists	1,513	3.91	1.147	-0.886
11	The standards (e.g., European Federation of Psychologists Association [EFPA], American Psychological Association [APA]) defining the minimum technical qualities of a test should be enforceable	1,500	3.92	0.873	-0.410
12	Legislation is needed to control the more serious abuses of testing	1,509	3.72	0.990	-0.451
14	Anyone who can demonstrate their competence as a test user (whether a psychologist or not) should be allowed to use tests	1,512	2.80	1.247	0.121
16	Controls on tests and testing should be minimal, as controls discourage the development of new ideas and new procedures	1,506	2.18	0.964	0.410
18	Publishers should be allowed to sell whatever tests they think fit	1,501	2.27	1.051	0.480

No.	Question	N	Mean	SD	Skew
19	Our National Psychological Association should take a more active role in the regulation and improvement of test use	1,501	4.12	0.852	-0.766
Computers and Internet					
5	In my professional field computer-based testing is progressively replacing paper and pencil tests	1,503	2.90	1.232	0.060
7	Test administration over the Internet has many advantages compared with paper-and-pencil administration	1,514	3.36	1.103	-0.240
10	Computer-generated interpretive reports do not have any validity	1,511	2.93	0.883	0.083
13	Test administration over the Internet sets some test takers at a disadvantage	1,505	3.35	1.044	-0.084
15	If properly managed, the Internet can greatly improve the quality of test administration	1,512	3.33	0.993	-0.236
17	The privacy of the test taker is not protected when testing by Internet	1,500	2.64	1.010	0.137
20	Testing over the Internet opens the way to fraud	1,496	3.36	0.988	-0.064
Test use					
21	I use tests regularly in the exercise of my profession	1,501	4.02	1.288	-1.146
22	Tests constitute an excellent source of information if they are combined and complemented with other psychological data	1,503	4.64	0.655	-2.118
23	Used correctly, tests are of great help to the psychologist	1,502	4.62	0.670	-1.990
24	All things considered, in the last decade, tests and testing practices have improved in my country	1,499	3.46	0.824	0.224
25a	Frequency with which you believe the following test-use problems occur: Making photocopies of copyrighted materials	1,458	3.70	1.200	-0.642
25b	Frequency with which you believe the following test-use problems occur: Making evaluations using inappropriate tests	1,456	2.98	0.978	-0.014
25c	Frequency with which you believe the following test-use problems occur: Not keeping up with the field	1,462	3.46	0.951	-0.271
25d	Frequency with which you believe the following test-use problems occur: Failing to check one's own interpretations with others	1,464	3.58	0.956	-0.340
25e	Frequency with which you believe the following test-use problems occur: Not considering errors of measurement of a test score	1,461	3.36	0.937	-0.192
25f	Frequency with which you believe the following test-use problems occur: Not restricting test administration to qualified personnel	1,456	3.11	1.142	-0.097
25g	Frequency with which you believe the following test-use problems occur: Not taking into account conditions that cast doubt on reported validity for a local situation	1,459	3.29	0.973	-0.155



No.	Question	N	Mean	SD	Skew
25h	Frequency with which you believe the following test-use problems occur: Making interpretations which go beyond the limits of the test	1,459	3.47	0.975	-0.257

Note. *N* = Number of responses to each item, *M* = Mean score, *SD* = Standard deviation, *Skew* = Skewness. Information and training

Professionals score low on whether they are provided with sufficient information regarding test quality (Item 4, $M = 2.43$, $SD = .94$) and feel they are not keeping pace with developments in the field (Item 25c, $M = 3.46$, $SD = .95$). The training received for the correct use of most tests during the final stage of professional studies was rated as average (Item 2, $M = 3.08$, $SD = 1.00$), a sentiment that was more pronounced among younger practitioners ($r = -.16$). The scores also indicate that respondents do not solely rely on the knowledge gained during their studies (Item 6, $M = 2.47$, $SD = 1.24$), with this trend being more evident in older practitioners ($r = -.19$).

Test Quality and Qualifications

Respondent ratings are high regarding their concerns about improper test use, particularly with respect to interpretation checks (Item 25d, $M = 3.58$, $SD = .96$), measurement errors (Item 25e, $M = 3.36$, $SD = .94$), validity limitations (Item 25g, $M = 3.29$, $SD = .97$) and other limitations (Item 25h, $M = 3.47$, $SD = .98$). Except for concerns related to illegal copying ($F = 17,350$, $df = 3$, $p < .001$), work psychologists generally expressed more concern than clinical psychologist about test use problems.

Regulation

Respondents scored above average regarding attitudes towards restricting psychological testing to qualified psychologists (Item 8, $M = 3.89$, $SD = 1.06$), particularly concerning the interpretation of the tests (Item 9, $M = 3.91$, $SD = 1.15$). This rating was higher among older psychologists compared to younger psychologists ($r = .23$ and $r = .11$ for Items 8 and 9, respectively), and higher among clinical psychologists than among work and educational psychologists ($F = 3,128$, $df = 3$, $p < .001$, and $F = 11,485$, $df = 3$, $p < .001$, for Items 8 and 9, respectively). However, respondents expressed mixed opinions on whether individuals who can demonstrate test user competence should be allowed to use tests (Item 14, $M = 2.80$, $SD = 1.25$). Older psychologists rated this item lower than younger psychologists ($r = -.19$). In terms of regulations, respondents clearly favoured national and European

standards, as well as legislative measures to regulate test use and abuse (Item 19, $M = 4.12$, $SD = .85$; Item 11, $M = 3.92$, $SD = .87$; and Item 12, $M = 3.72$, $SD = .99$; respectively).



Computers and internet-based testing

Internet-based testing was considered by respondents to improve the quality of test administration if properly managed (Item 15, $M = 3.33$, $SD = .99$). A high mean score (Item 13, $M = 3.35$, $SD = 1.04$) suggests that psychologists perceive potential disadvantages of internet-based testing for test takers, yet they generally report more positive attitudes towards its overall potential and possibilities (Item 7, $M = 3.36$, $SD = 1.10$). In this regard, age played an important role, as older psychologists rated internet-based testing less positively than younger psychologists (r 's from .14 to .22). Work psychologists more frequently reported that computer-based testing is replacing pen-and-paper testing (Item 5, $F = 35,056$, $df = 3$, $p < .001$), and they report more positive attitudes towards internet-based testing (Item 7, $F = 15,576$, $df = 3$, $p < .001$; Item 13, $F = 7,526$, $df = 3$, $p < .001$; Item 15, $F = 9,820$, $df = 3$, $p < .001$).

Test use

The respondents listed a total of 170 different tests among their top 3 most frequently used (the exact count depended on how we grouped them). Table 2 shows the most frequently used tests by count and percentage. The *Wechsler tests* (Wechsler, 1987, 2008, 2014a, 2014b; Wechsler & Naglieri, 2006) were by far the most frequently used at 12.8%, followed by various *SCID* versions (First et al., 1997, 2002, 2016) at 6.8%. The new Norwegian translation of the *SCID-5-CV* (First et al., 2018) had increased in usage but only represented 0.9% of the total. Most psychologists reported using tests regularly (Item 21, $M = 4.02$, $SD = 1.29$), and they were considered an excellent source of information (Item 22, $M = 4.64$, $SD = .66$) and to be of great help (Item 23, $M = 4.62$, $SD = 0.67$).

Table 2

Test Use Frequencies and Percentages

No.	Short Name	Frequency	
1	Wechsler Tests	195	12.8%
2	SCID	104	6.8%
3	MINI	61	4.0%
4	CONNERS	57	3.7%

No.	Short Name	Frequency	
5	D-KEFS	38	2.5%
6	BDI	36	2.4%
7	SCL-90	35	2.3%
8	MMPI	33	2.2%
9	BRIEF	25	1.6%
10	BAI	23	1.5%
11	CVLT	23	1.5%
12	DIVA	19	1.2%
13	ADOS	16	1.1%
14	MADRS	15	1.0%
15	NEPSY	15	1.0%
16	AUDIT	14	0.9%
17	DUDIT	14	0.9%
18	ASEBA	13	0.9%
19	BAYLEY	13	0.9%
20	LEITER	13	0.9%
21	VINELAND	13	0.9%
22	CELF	12	0.8%
23	DES	12	0.8%
24	PANSS	12	0.8%
25	IIP	9	0.6%
26	TMT	9	0.6%
27	RCFT	8	0.5%
28	Rorschach	8	0.5%
29	WMS	7	0.5%
30	Halstead-Reitan	6	0.4%



A priori subscales

Internal consistency varied considerably across the five subscales. The *Concern* subscale had high reliability ($\alpha = .854$), whereas the *Regulations* and *Internet* subscales exhibited very low reliability ($\alpha = .169$ and $.159$, respectively). The *Appreciation* and *Knowledge* subscales showed moderate consistency, with Cronbach's alphas of $.643$ and $.586$, respectively.

As shown in Table 3, there are significant correlations between *Concern*, *Regulations* and *Internet*, but not for *Appreciation* and *Knowledge*. Further, *Regulations* correlates with all factors except *Knowledge*, whereas *Internet* correlates with all but *Appreciation*.



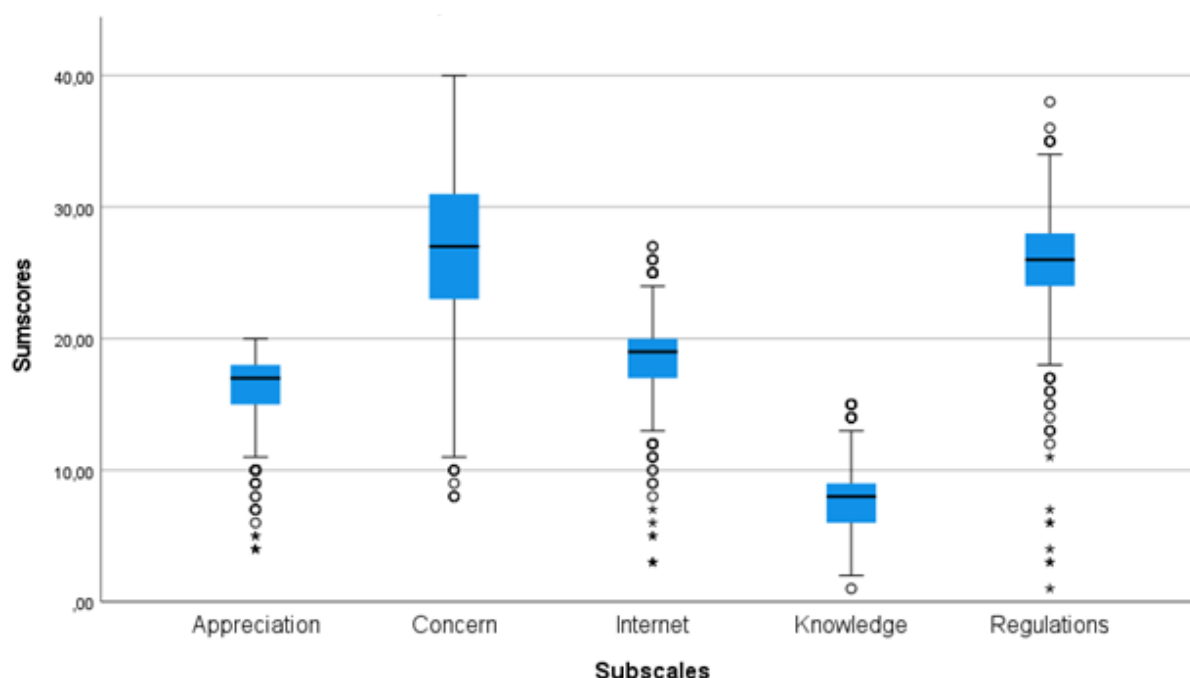
Table 3

Correlations Between A Priori Sub Scales

	Regulation	Internet	Appreciation	Knowledge
Concern	.183**	.153**	-.020	.042
Regulations		.300**	.170**	.041
Internet			.003	.053*
Appreciation				-.156**

* $p < .05$ ** $p < .005$

Since several of the Evers factors showed poor internal consistency in our data, we present a boxplot in Figure 1 for the visual presentation of the different factors. The whiskers overlap considerably, suggesting a high degree of covariation. Moreover, numerous extreme studentised outliers (Grubbs, 1969) – those exceeding twice the interquartile range – along with a number of outliers beyond 1.5 times the interquartile range – significantly affect the *Appreciation*, *Internet* and *Regulations* factors. The factors therefore seem neither sound nor useful.

Figure 1*A Priori Factors Boxplot*

Note. X Axis Subscales: Appreciation, Concern, Internet, Knowledge, Regulations; Y Axis Sum Scores: 0, 10, 20, 30, 40.

Discussion

The main aim of this study was to report on attitudes towards and concerns about test use and test quality from a survey among Norwegian psychologists. Our primary finding is that test quality and regulations are major concerns, as indicated by high ratings on items relating to the need to document test quality, concerns about improper test use and a clear endorsement of more restrictions on who may use tests. This result is reflected in various aspects of the results and echoes the concerns raised in a previous Norwegian study on test use (Vaskinn et al., 2010). They reported that 96% of respondents were ‘somewhat’ or ‘very’ concerned about test quality, a sentiment that was also communicated in our survey. They also revealed differences in the degree to which practitioners checked for quality themselves, whereas our study findings could be interpreted as a call for more assistance in quality checking.

Our respondents reported satisfaction with the training they received in their professional psychology studies but also indicated that they acquired more knowledge after graduating. The *Board of Test Policy* recently conducted an informal review of university assessment training, and together



with the current survey, we find support for our view that training at the basic level seems adequate. Given that challenges related to learning specific tests and acquiring practical assessment skills are addressed (Handler & Meyer, 1998), efforts should be directed more specifically towards providing easily accessible information on test quality to enable clinicians to make informed choices.

In Norway, the health authorities have explicitly delegated the responsibility for test quality evaluation to the individual clinician. However, Vaskinn and Egeland (2012) demonstrated that only half of their respondents checked test quality themselves. The *Norwegian Psychologists Association* has challenged this decision, arguing that it imposes an excessive demand on clinicians, given the workload and expertise required to evaluate a test thoroughly. The available documentation for many tests is often rudimentary, at times inaccurate (Mihura et al., 2019) or even non-existent, which is also a concern raised by our respondents. Some institutions have made efforts towards systematically evaluating the quality of commonly used test methods, but sustaining this work over time has proven difficult. Additionally, as noted by a reviewer of this manuscript, in a small country like Norway, we have a limited selection of translated tests from which to choose, rendering us more reliant on the quality of the few tests available. This highlights the need for more coordinated efforts to regulate and guide the field.

The lack of effort from clinicians to systematically evaluate test quality is not due to a shortage of engagement or competence; on the contrary, clinicians report that they are invested in assessment and recognise its value. Nevertheless, the challenging and comprehensive task of evaluating test quality requires time and effort with resources that can only be allocated through national initiatives. We therefore find it unreasonable that the individual clinician should bear this responsibility alone. Our current data indicate that the respondents support our view. Test quality and competent use are major concerns for the respondents, who appear frustrated with the current situation.

Several survey items show that our respondents are generally positive about computer- and internet-based testing, particularly among the younger generation. Interest in this issue may have waned somewhat. In clinical practice, paper testing remains widespread. Several factors may account for this, including the lack of appropriate licensing schemes that limit access to computer- and internet-based systems in hospitals and clinics; concerns over data privacy due to the possibility of data being stored on foreign servers; and cost. Practitioners may also be sceptical because they have limited oversight over the conditions under which patients complete tests at home or in other



locations. This scepticism appears to be more pronounced among older practitioners, which may be expected. The lower level of concern about illegal copying within the field of work psychology may stem from the reduced prevalence of paper-and-pencil testing, as this is a field that has adopted computer-based testing to a greater extent.

The question of test use restrictions is a more contentious issue. Respondents expressed clear support for greater national and international regulation in this area, preferably through legislative measures. They are more divided on whether other professions and groups should be permitted to use tests based on demonstrated qualifications. There seems to be a clear sentiment towards restricting test interpretation to licensed psychologists. However, this remains a contentious issue among our respondents, reflecting tensions between those advocating for restrictions based on specific educational backgrounds and those supporting access for individuals with proven qualifications. Notably, older and clinical practitioners tend to be more restrictive in this regard, while younger practitioners in work and educational sectors are less so.

The usefulness of grouping the questionnaire items into sub-factors appears to be quite limited in our data. Only the factor *Concern* achieves an acceptable level of internal consistency; the other two factors are critically low. In addition, the factors correlate to a high degree. We therefore suggest that future analyses of the EFPA survey should focus solely on the individual items – or, alternatively, that analyses at the sub-factor level be conducted on the international data set as a whole.

A major limitation of the test use section of the survey is its restriction to the three most commonly used tests. This means that we do not have a comprehensive test use survey – only indications of test use. With this in mind, our results indicate that performance-based cognitive tests are the most frequently used, even though neuropsychologists constitute a small portion of the sample. Conversely, clinically oriented performance-based tests, such as the Rorschach Inkblot Method, are used relatively infrequently. We can only speculate that users of the latter may not consider themselves to be test-savvy, contributing to their underrepresentation in the current data, as psychologists in private practice also constituted a small subgroup of the sample. The only broad personality and psychopathology test that ranks highly on the list is the MMPI (Butcher et al., 2011). It is also surprising to see the continued use of the SCL-90 (Derogatis, 1971), despite its well-known, serious psychometric limitations (Groth-Marnat & Wright, 2016). Strictly speaking, the SCID, MINI (Sheehan et al., 2009), MADRS (Montgomery & Åsberg, 1979), AUDIT (Babor et al., 1992) and

DUDIT (Berman et al., 2007) can be described as checklists rather than psychometric tests, but methods like these nonetheless provide structure to assessments.



Earlier data from the United States show that self-report methods are by far the most popular type used by psychologists (Archer et al., 2006; Norcross & Karpiak, 2012), followed by performance-based methods. All practitioners involved in emotional injury cases who responded to a survey (Boccaccini & Brodsky, 1999) reported using the MMPI-2, with 54% using the Wechsler tests, 50% using the MCMI-III (Millon Clinical Multiaxial Inventory-III; Millon, 1994) and 41% using the Rorschach Inkblot Method (Rorschach, 1921). Data from as far back as 1985 indicate a decline in the use of performance-based methods, such as the Rorschach and the Wechsler tests (Piotrowski, 2015; Piotrowski et al., 1985). This trend corresponds with the increased research on and use of self-report tests in recent decades (Baumeister et al., 2007). Nonetheless, there are also indications that the use of tests is stable. Wright et al. (2017) demonstrated that self-report tests dominate among practitioners, alongside ability tests. In their survey, the Rorschach Inkblot Method and other performance-based ('projective') methods ranked eighth and ninth in terms of usage.

More generally, variations in test usage are also evident across different parts of the world. Some methods have a relatively strong foothold in certain countries but not in others. We know first-hand that the Rorschach Inkblot Method has very limited use in, for example, Great Britain and Germany, whereas it is widely employed in Finland and the United States. Another example is the *Wartegg Drawing Completion Test*, which is frequently used by psychologists in Sweden and Finland but is practically unknown in neighbouring Norway (Soilevuo Grønnerød & Grønnerød, 2012). In a small country like Norway, test education and usage are more dependent on specific professionals who actively teach and promote the use of particular tests. This situation disproportionately influences which tests are used compared to larger countries. That said, studies also indicate that most tests in use are common across countries, at least in the Western sphere (Egeland et al., 2016; Muñoz et al., 2001).

The main limitation of our study is the response rate of about 20%. Surveys of this nature have generally concluded that European psychologists have a positive attitude towards testing. However, generalisations must be made cautiously, as the response rates are generally low (15–20% across most participating countries). Even if the dispersion across work fields and specialties is representative of the psychologist population, the potential for self-selection is apparent. We must therefore assume that those most invested in tests and assessments are also those most likely to respond to a questionnaire of this kind. We can at best presume that the results are representative



of assessment-focused psychologists, and less so for the psychologist population at large. Another limitation is that respondents were asked to name the three most-used tests in their practice. A more complete list of tests would have been useful, as it would have revealed differences in test diversity among practitioners and provided a more comprehensive overview of the most commonly used tests.

Conclusions

In the Norwegian segment of the EFPA's 2020 decennial survey of psychologists' attitudes towards tests and test use, test quality and regulation are the primary concerns expressed by respondents. Despite a relatively low response rate of 20%, study results highlight the need for more accessible information on test quality and for increased regulation based on user competence.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A Survey of Psychological Test Use Patterns Among Forensic Psychologists. *Journal of Personality Assessment*, 87(1), 84–94. https://doi.org/10.1207/s15327752jpa8701_07
- Babor, T. F., Fuente, J. R. D. L., Saunders, J., Grant, M., Babor, T. F., Fuente, J. R. D. L., Saunders, J., Grant, M., & T. H. (1992). *AUDIT - The Alcohol Use Disorders Identification Test: Guidelines for use in Primary Health Care*. World Health Organization.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on Psychological Science*, 2(4), 396–403. <https://doi.org/10/c5f>
- Berman, A. H., Bergman, H., Palmstierna, T., & Schlyter, F. (2007). *DUDIT –The Drug Use Disorders Identification Test, MANUAL Version 1.1*. Karolinska Institute, Department of Clinical Neuroscience, Section for Alcohol and Drug Dependence Research.
- Boccaccini, M. T., & Brodsky, S. L. (1999). Diagnostic Test Usage by Forensic Psychologists in Emotional Injury Cases. *Professional Psychology: Research and Practice*, 30(3), 253–259. <https://doi.org/10/dm7sz8>



- Butcher, J. N., Beutler, L. E., Harwood, T. M., & Blau, K. (2011). The MMPI-2. In T. M. Harwood, L. E. Beutler, & G. Groth-Marnat (Eds.), *Integrative Assessment of Adult Personality* (3rd Edition, pp. 152–189). Guilford Press.
- Derogatis, L. R. (1971). *Symptom Checklist-90 – Revised*. <https://doi.org/10.1037/t01210-000>
- Egeland, J., Løvstad, M., Norup, A., Nybo, T., Persson, B. A., Rivera, D. F., Schanke, A.-K., Sigurdardottir, S., & Arango-Lasprilla, J. C. (2016). Following international trends while subject to past traditions: Neuropsychological test use in the Nordic countries. *The Clinical Neuropsychologist*, 30(sup1), 1479–1500. <https://doi.org/10.1080/13854046.2016.12376755>
- European Federation of Psychologists' Associations. (2023, 23 June). *Ongoing revision of the European Test Review Model | EFPA*. <https://www.efpa.eu/ongoing-revision-european-test-review-model>
- Evers, A. (2012). The Internationalization of Test Reviewing: Trends, Differences, and Results. *International Journal of Testing*, 12(2), 136–156. <https://doi.org/10/gntfsb>
- Evers, A., McCormick, C. M., Hawley, L. R., Muñiz, J., Balboni, G., Bartram, D., Boben, D., Egeland, J., El-Hassan, K., Fernández-Hermida, J. R., Fine, S., Frans, Ö., Gintiliené, G., Hagemester, C., Halama, P., Iliescu, D., Jaworowska, A., Jiménez, P., Manthouli, M., ... Zhang, J. (2017). Testing Practices and Attitudes Toward Tests and Testing: An International Survey. *International Journal of Testing*, 17(2), 158–190. <https://doi.org/10/gbv2pq>
- Evers, A., Muñiz, J., Hagemester, C., Høstmælingen, A., Lindley, P., Sjöberg, A., & Bartram, D. (2013). Assessing the quality of tests: Revision of the EFPA review model. *Psicothema*, 25(3), 283–291. <https://doi.org/10.7334/psicothema2013.97>
- First, M. B., Gibbon, M., Spitzer, R. L., Williams, J. B. W., & Benjamin, L. S. (1997). *Structured clinical interview for DSM-IV axis II personality disorders, (SCID-II)*. Washington, DC: American Psychiatric Association Publications.
- First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders, research version, patient edition (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2016). *Structured Clinical Interview for DSM-5 Disorders, Clinician Version (SCID-5-CV)*. Arlington, VA, American Psychiatric Association.



- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2018). *Strukturert klinisk intervju for psykiske lidelser—DSM-5*. [Structured Clinical Interview for Mental Disorders—DSM-5] Gyldendal.
- Greiff, S. (2017). The Field of Psychological Assessment: Where it Stands and Where it's Going – A Personal Analysis of Foci, Gaps, and Implications for EJPA. *European Journal of Psychological Assessment*, 33(1), 1–4. <https://doi.org/10.1027/1015-5759/a000412>
- Groth-Marnat, G., & Wright, A. J. (2016). *Handbook of Psychological Assessment*. John Wiley & Sons.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12(1), 19–30. <https://doi.org/10/dr95m>
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1), 1–21. <https://doi.org/10.2307/1266761>
- Grønnerød, C. (2024). *Psykologisk utredning av voksne. En teoretisk og praktisk innføring*. [Psychological assessment of adults. A theoretical and practical introduction.] Cappelen Damm Akademisk.
- Handler, L., & Meyer, G. J. (1998). The Importance of Teaching and Learning Personality Assessment. In L. Handler & M. J. Hilsenroth (Eds.), *Teaching and Learning Personality Assessment* (pp. 3–30). Routledge.
- International Test Commission. (2001). International Guidelines for Test Use. *International Journal of Testing*, 1(2), 93–114. https://doi.org/10.1207/S15327574IJT0102_1
- International Test Commission. (2006). International Guidelines on Computer-Based and Internet-Delivered Testing: International Journal of Testing: Vol 6, No 2. *International Journal of Testing*, 6(2), 143–171. https://doi.org/10.1207/s15327574ijt0602_4
- Kornør, H. (2015). Hvor ofte tenker du på om testen du bruker, er psykometrisk god nok? [How often do you consider whether the test you use is psychometrically good enough?] *Tidsskrift for Norsk Psykologforening*, 52(7), 600–601.
- Krishnamurthy, R., Natoli, A. P., Arbisi, P. A., Hass, G. A., & Gottfried, E. D. (2022). Professional Practice Guidelines for Personality Assessment: Response to Comments by Ben-Porath (2022), Lui (2022), and Jenkins (2022). *Journal of Personality Assessment*, 104(1), 27–29. <https://doi.org/10/gnxbjw>



- Mihura, J. L., Bombel, G., Dumitrascu, N., Roy, M., & Meadows, E. A. (2019). Why We Need a Formal Systematic Approach to Validating Psychological Tests: The Case of the Rorschach Comprehensive System. *Journal of Personality Assessment*, 101(4), 374–392. <https://doi.org/10/gntfpj>
- Millon, T. (1994). *Millon Clinical Multiaxial Inventory-III: Manual for the MCMI-III*. Pearson Assessment, Inc.
- Montgomery, S. A., & Åsberg, M. (1979). A New Depression Scale Designed to be Sensitive to Change. *The British Journal of Psychiatry*, 134(4), 382–389. <https://doi.org/10.1192/bjp.134.4.382>
- Muñiz, J., Bartram, D., Evers, A., Boben, D., Matesic, K., Glabeke, K., Fernández-Hermida, J. R., & Zaal, J. N. (2001). Testing practices in European countries. *European Journal of Psychological Assessment*, 17(3), 201–211. <https://doi.org/10.1027/1015-5759.17.3.201>
- Norcross, J. C., & Karpiak, C. P. (2012). Clinical Psychologists in the 2010s: 50 Years of the APA Division of Clinical Psychology. *Clinical Psychology: Science and Practice*, 19(1), 1–12. <https://doi.org/10/gntfsf>
- Piotrowski, C. (2015). Projective Techniques Usage Worldwide: A review of applied settings 1995-2015. *Journal of the Indian Academy of Applied Psychology*, 41(3), 9–19.
- Piotrowski, C., Sherry, D., & Keller, J. W. (1985). Psychodiagnostic test usage: A survey of the society for personality assessment. *Journal of Personality Assessment*, 49(2), 115–119. <https://doi.org/10/ckdv3d>
- Rorschach, H. (1921). *Psychodiagnostik*. [Psychodiagnostics.] Hans Huber Verlag.
- Ryder, T. (2021a). Testkvalitetsprosjektet - del 1: Norske psykologers testholdninger og testbruk. [The Test Quality Project – part 1: Norwegian psychologists' test attitudes and test use.] *Tidsskrift for Norsk psykologforening*, 58(1), 28–37.
- Ryder, T. (2021b). Testkvalitetsprosjektet – del 2: Tester med behov for kvalitetstiltak. [The Test Quality Project – part 2: Tests in need of quality improvements.] *Tidsskrift for Norsk psykologforening*, 58(2), 92–105.
- Sheehan, D., Janavs, J., Harnett-Sheehan, K., Sheehan, M., & Gray, C. (2009). *Mini International Neuropsychiatric Interview 6.0.0*. Mapi Research Institute.



- Soilevuo Grønnerød, J., & Grønnerød, C. (2012). The Wartegg Zeichen Test: A Literature Overview and a Meta-Analysis of Reliability and Validity. *Psychological Assessment*, 24(2), 476–489. <https://doi.org/10/dhf7gp>
- Vaskinn, A., & Egeland, J. (2012). Testbruksundersøkelsen: En oversikt over tester brukt av norske psykologer. [The test user survey: An overview of tests used by Norwegian psychologists.] *Tidsskrift for Norsk psykologforening*, 49(7), 658–665.
- Vaskinn, A., Egeland, J., Høstmark Nielsen, G., & Høstmælingen, A. (2010). Norske psykologers bruk av tester. [Norwegian psychologists' test use.] *Tidsskrift for Norsk psykologforening*, 47(11), 1010–1016.
- Wechsler, D. (1987). *WMS-R: Wechsler Memory Scale--Revised : Manual*. Psychological Corporation.
- Wechsler, D. (2008). *Wechsler Adult Intelligence Scale: WAIS-IV ; Technical and Interpretive Manual*. Pearson.
- Wechsler, D. (2014a). *WISC-V: Administration and Scoring Manual*. PsychCorp.
- Wechsler, D. (2014b). *WPPSI-IV: Wechsler Preschool and Primary Scale of Intelligence - fourth edition : Manual*. Pearson Education.
- Wechsler, D., & Naglieri, J. A. (2006). *WNV: Wechsler Nonverbal Scale of Ability ; Administration and Scoring Manual*. PsychCorp.
- Wright, C. V., Beattie, S. G., Galper, D. I., Church, A. S., Bufka, L. F., Brabender, V. M., & Smith, B. L. (2017). Assessment practices of professional psychologists: Results of a national survey. *Professional Psychology: Research and Practice*, 48(2), 73–78. <https://doi.org/10.1037/pro0000086>
- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S., Nichols, C. N., Lampropoulos, G. K., Walker, B. S., Cohen, G., & Rush, J. D. (2006). The Meta-Analysis of Clinical Judgment Project: Fifty-Six Years of Accumulated Research on Clinical Versus Statistical Prediction. *The Counseling Psychologist*, 34(3), 341–382. <https://doi.org/10/cbrmjf>