

# Considerations for a policy of best practices in the use of psychological tests in Chile

## *(Consideraciones para una política de buenas prácticas en el uso de pruebas psicológicas en Chile)*

Marcela Rodríguez-Cancino

Department of Psychology, Universidad de La Frontera, Chile; marcela.rodriguez@ufrontera.cl 

**Reference:** Rodríguez-Cancino, M. (2025). Considerations for a policy of best practices in the use of psychological tests in Chile (*Consideraciones para una política de buenas prácticas en el uso de pruebas psicológicas en Chile*). *Ibero-American Journal of Psychology and Public Policy*, 2(1), 11-46. <https://doi.org/10.56754/2810-6598.2025.0029>

**Editor:** Rodrigo Ferrer

**Reception date:** 13 Sept 2024

**Acceptance date:** 26 Nov 2024

**Publication date:** 28 Jan 2025

**Language:** English and Spanish

**Translation:** Helen Lowry

**Publisher's Note:** IJP&PP remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2025 by the author. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY NC SA) license (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

**Abstract:** Several authors have mentioned that best practices in using tests require that the psychologist has solid training in the area, that the quality of the instruments is guaranteed, and that they are used ethically and rigorously. These aspects mainly challenge training for professional practice, research, and regulation. Developing skills for psychological assessment at the undergraduate level, along with psychometric research, has posed a major challenge in Chile, as has the concern for the adequate regulation of test use. This is an important issue, as the tests, in conjunction with other data, support decision-making in various areas of the specialized discipline of psychology, affecting many people's lives. This paper briefly reviews the main international guidelines that ensure the appropriate use of tests and refers to these issues. Competent and qualified professionals must administer tests; there must be ongoing assessment of their psychometric quality through research into validity, reliability, and fairness testing; and there must be mechanisms in place to ensure rigor in the selection and administration of tests by psychologists. The importance and necessity of articulating these elements in creating a national policy that safeguards the rights of all test-takers is discussed.

**Keywords:** psychological assessment; professional competence; training; psychometric properties; regulation; professional practice.

**Resumen:** Diversos autores han mencionado que las buenas prácticas en el uso de test requieren que el/la psicólogo/a cuente con una sólida formación en el área, que la calidad de los instrumentos esté garantizada y que se utilicen de manera ética y rigurosa. Estos aspectos interpelan principalmente a la formación, investigación y regulación del ejercicio profesional. En Chile, se han evidenciado grandes desafíos en el desarrollo de competencias para la evaluación psicológica a nivel de pregrado y la investigación en psicometría, así como también importantes falencias en el adecuado control del uso de instrumentos. Esto es un problema importante ya que los test, junto a otras fuentes de información, fundamentan la toma de decisiones en distintos contextos del ejercicio profesional de la psicología, afectando la vida de muchas personas. En el presente documento se aborda una breve revisión de las principales directrices internacionales que garantizan el buen uso de test, y que se refieren a estos temas. Se resalta la importancia de que la administración de los test esté a cargo de profesionales competentes y calificados/as; que exista una valoración permanente de la calidad psicométrica de los instrumentos a través de la exploración de las evidencias validez, confiabilidad e imparcialidad; y que existan mecanismos que garanticen la rigurosidad con que los/las psicólogos/as seleccionan y administran test, o realizan la interpretación y comunicación de sus resultados. Se discute la importancia y necesidad de articular estos elementos en la creación de una política nacional que resguarde los derechos de todas las personas evaluadas.

**Palabras clave:** evaluación psicológica; competencia profesional; formación; propiedades psicométricas; regulación; ejercicio profesional.

**Resumo:** Vários autores referiram que as boas práticas na utilização de testes exigem que o psicólogo tenha uma sólida formação na área, que a qualidade dos instrumentos seja garantida e que sejam utilizados de forma

ética e rigorosa. Estes aspectos dizem sobretudo respeito à formação, à investigação e à regulamentação do exercício profissional. No Chile, têm sido evidentes grandes desafios no desenvolvimento de competências para a avaliação psicológica na graduação e na investigação em psicometria, bem como importantes deficiências no controlo adequado da utilização dos instrumentos. Este é um problema importante, uma vez que os testes, juntamente com outras fontes de informação, apoiam a tomada de decisões em diferentes contextos da prática profissional da psicologia, afetando a vida de muitas pessoas. Este documento aborda uma breve revisão das principais orientações internacionais que garantem a utilização adequada dos testes e que se referem a estes temas. Salienta-se a importância de a aplicação dos testes ser realizada por profissionais competentes e qualificados; que haja uma avaliação permanente da qualidade psicométrica dos instrumentos através da exploração de evidências de validade, fiabilidade e imparcialidade; e que existem mecanismos que garantem o rigor com que os psicólogos seleccionam e administram os testes, ou interpretam e comunicam os seus resultados. Discute-se a importância e a necessidade de articular estes elementos na criação de uma política nacional que proteja os direitos de todas as pessoas avaliadas.

*Palavras-chave:* avaliação psicológica; competência profissional; formação; propriedades psicométricas; regulamento; prática profissional.

## 1. Introduction

The term "assessment" denotes the intricate process of synthesizing knowledge, clinical judgment, various information sources, and psychometric constructs within a problem-solving framework (American Psychological Association [APA], 2020). Psychological assessment is a process with specifically defined stages or phases that guide decision-making through scientific and professional actions for collecting, assessing, and integrating information about a person. This procedure is executed based on action planning, where the use of diverse information sources is crucial (Fernández-Ballesteros et al., 2003; Fernández-Ballesteros et al., 2011; Hernández et al., 2021; Marín, 2021).

These information sources include tests, defined as any measuring device or instrument that evaluates psychological constructs through a person's responses to a series of items. Tests seek to capture a structured sample of a person's behavior in a specific domain to guide decision-making in various fields of psychology (APA, 2020; Ferrando et al., 2022; Muñoz et al., 2020). Tests can be central in a psychological assessment process; however, they must be used correctly to function as reliable tools for professional practice (Prieto & Muñoz, 2000).

The appropriate use of psychological tests is safeguarded by international entities such as the American Education Research Association (AERA), the International Test Commission (ITC), and the APA, which widely recognize their scientific contribution to the area. These organizations have been engaged in achieving consensus about, generating, and implementing measures and guidelines in this area (Díaz-López et al., 2017; Hernández et al., 2021).

### 1.1 Foundations for the proper use of tests in the Chilean setting

Several authors agree that the usefulness of tests lies in the proper articulation of three essential pillars (Elosua, 2017; Evers et al., 2012; Hernández et al., 2021; Hernández et al., 2022; Muñoz et al., 2011; Muñoz et al., 2020):

- 1) competent and qualified professionals should administer the tests,
- 2) the quality (psychometric properties) of the tests must be guaranteed, and
- 3) the use of the tests must adhere to ethical standards and rigor.

The first of these pillars refers essentially to the field of professional training. The second is a task performed mainly by research teams in the area. The third refers to an area that should be safeguarded through adequate mechanisms for regulating professional practice.

The current scenario in Chile shows great challenges concerning these three pillars. Concerning professional training, Vinet and Rodríguez-Cancino (2024) point out that in recent decades, there has been a significant increase in undergraduate psychology programs, with little formalization and/or without minimum quality assurance mechanisms. The authors mention that there have been efforts to ensure the quality of training, such as the creation of the Network of Schools of Psychology of the Consortium of State Universities of Chile, which has established that psychological assessment is a central macro-competency for undergraduates. However, on the one hand, there is still no total agreement or regulation on the contents of psychological and psychometric assessment that should be taught. On the other, the study by Vinet et al. (2023) reports that the psychologists surveyed consider that initial training has not been sufficient for the adequate use of tests in the professional world. Accordingly, Chile requires that the professional training in this area be enhanced, giving it a specialized and/or graduate nature.

Concerning the second pillar, Vinet and González (2013) observe that deficiencies in training in Chile have led to a reduction in the creation or adaptation of tests, which a rise in the number of graduate programs has offset. According to Vinet et al. (2023), instrument research and development in Chile was enriched after the creation of national institutions charged with these tasks; however, it is imperative to bolster research teams focused on the construction, translation, and/or adaptation of instruments, rigorously applying international guidelines. In addition, they point out that mechanisms must be implemented to assess the tests' psychometric quality, standards, and cultural relevance.

Finally, the third pillar refers to the need for mechanisms to regulate professional practice. In the 1980s, the Association of Psychologists in Chile (translated as Colegio de Psicólogas y Psicólogos de Chile) lost its legal guardianship over professional activity, resulting in an unregulated environment of unregulated practice (Vinet et al., 2023; Vinet & González, 2013). This circumstance has resulted in the widespread use of instruments lacking validation in their psychometric properties or evidence of their relevance to the Chilean population; hence, there must be strategies to regulate practices, including accreditation systems for test usage, guidelines about which tests can be used and which cannot, and penalties for non-compliance with minimum conditions for their use (Vinet & Rodríguez-Cancino, 2024; Wechsler et al., 2022).

These challenges in training, research, and regulation on the appropriate use of tests support the need to begin working on measures to ensure best practices in the psychological assessment processes in Chile.

### ***1.2 Why is there a need for a policy on best practices in test use?***

The individuals being assessed are entitled to the application of instruments that adhere to professional standards of technical quality, administered under appropriate conditions, with objectivity and fairness governing the interpretation and reporting of the findings. The incorrect application of tests in psychological assessment processes primarily infringes on these rights and seriously affects the people being assessed, as critical decisions regarding their lives depend on these assessments (AERA et al., 2018; Muñiz & Hambleton, 1996).

Research concerning the perception of issues related to test usage consistently identifies several prevalent concerns, including their use by unqualified people, the use of inappropriate instruments, insufficient understanding of psychometric quality, and the reproduction of copyrighted information. Foreign tests with poor adaptation processes or lack of validation studies have also been detected (Evers et al., 2012; Hernández et al., 2021; Muñiz et al., 2020; Prieto et al., 1999).

In Chile, the study by Cataldo et al. (2012) explored the psychometric quality of the Alertness Test. This instrument is widely used in recruitment processes to assess risk propensity in the mining sector. The report identifies several problems in its use, such as the lack of a definition of the psychological constructs it intends to assess, the lack of an administration manual, and that it is used mainly through photocopies without respecting copyrights. This test features items decontextualized from the relevant sector and lacks standardized guidelines for results interpretation; thus, each organization or professional arbitrarily determines a score for selection. The results of this study showed that the test has no evidence concerning its reliability or validity level, and the authors suggest discontinuing its use due to its inaccuracy and the risk of using it to estimate the possibility of accidents. They also suggest that psychologists should not remain passive towards tests, the origin and quality of which as an assessment instrument are unknown.

Another national study revealed the extensive use of tests and techniques for personnel selection that do not have a scientific basis to support their use and predictive ability. It should be noted that this study detected that employers routinely require these tests, requesting psychologists who are trained in them to apply them. The author of this study warns that the widespread use of this type of test and bad practices in this area are related to shortcomings in professional training and the lack of quality control mechanisms for this training, the absence of regulation of professional practice by organizations such as the Chilean Association of Psychologists, and few spaces for the dissemination of specialized research in the area (Didier, 2014).

On the other hand, the study by Rodríguez et al. (2017) establishes that in Chile, the deregulation of consulting firms and poor professional development have led to the emergence of bad practices, among which is the use of assessment instruments of low technical quality, i.e., without relevance, reliability, or scientific validity. The authors point out that bad practices in this context incur significant costs for both the organization and the applicants, as well as for the psychologists responsible for conducting the screening processes. On this last point, for psychologists, observing or participating in substandard practices heightens the likelihood of encountering numerous stressors, including moral distress, a phenomenon that has been scarcely examined in Chile.

More recently, the only study in Chile that has addressed psychologists' perception of the use of tests, conducted by Vinet et al. (2023), reports the habitual use of tests with no empirical basis and/or relevance for the national population. This study identified that 13 of the 20 most frequently used tests were projective instruments. Furthermore, a systematic review revealed that among the 20 most prevalent tests, only five had undergone examination regarding their psychometric quality within the national population, specifically concerning reliability, validity evidence, and normative data; all of these were psychometric tests, with none being projective.

In light of the national context and the significant repercussions of improper testing practices on people's lives, mechanisms in Chile must be established urgently to ensure the integrity of training and their appropriate application, alongside promoting research in psychometrics. These mechanisms can be integrated into a national policy that regulates the use of psychological tests, mitigates

malpractice under professional ethical standards, and guarantees the utmost respect for every person assessed. Creating clear and effective regulatory mechanisms could guarantee greater quality, equity, objectivity, and safety in the psychological assessment process in Chile.

## 2. Objective

This document presents a review of the main guidelines that have been deployed internationally on the proper use of psychological assessment instruments, which can be linked to the three main pillars mentioned above, to identify the essential elements required to begin to develop a policy of best practices in test use in Chile.

## 3. Method

A descriptive narrative literature review was conducted. This methodology consists of presenting an updated synthesis of various data sources on a specific research topic, accompanied by the comments, recommendations, or conclusions of the authors of the review based on their experience and knowledge (Ato et al., 2013; Reyes, 2020; Pardal-Refoyo, 2023; Vera, 2009). The data sources selected for this review include scientific articles and instructions or manuals on specific technical guidelines on the proper application of tests prepared by organizations specializing in measurement and assessment in psychology with international experience, recognition, and validation.

First, a search strategy was defined in the most relevant electronic databases in the Social Sciences, such as Web of Science, Scopus, and Google Scholar. The SciELO database, APA and ITC websites, books, manuals, and technical guides of international organizations specialized in the subject were also included. The keywords that guided the search for scientific articles were “test”, “test use”, “guidelines”, “best practices”, and “professional practice” in Spanish. Second, the criteria for selecting data sources considered their relevance to the review objective (Guirao-Goris et al., 2008). Accordingly, sources were selected that (a) provide guidelines on the proper use of tests, (b) describe collective regulatory efforts on the use of tests performed and assessed in different countries, and (c) have recognition or evidence of consensus from the scientific community on the use of tests. According to the fulfillment of these criteria, the data sources selected to organize the proposed guidelines are presented in Table 1.

Finally, and in accordance with this methodology, following the search, review, and selection of sources, the information was classified into files, which serve as organizational systems that facilitate the ordering and prioritization of the collected data, determining its relevance and generating a sequential analysis for its eventual inclusion in the text composition (Loayza, 2021). Thus, this narrative literature review presents an updated synthesis of the main international guidelines specialized in the subject, including conclusions and recommendations by the author of this paper, to offer guidelines for the design of a best practices policy regarding the use of psychological tests.

**Table 1**

*Data sources used to organize the guideline proposal*

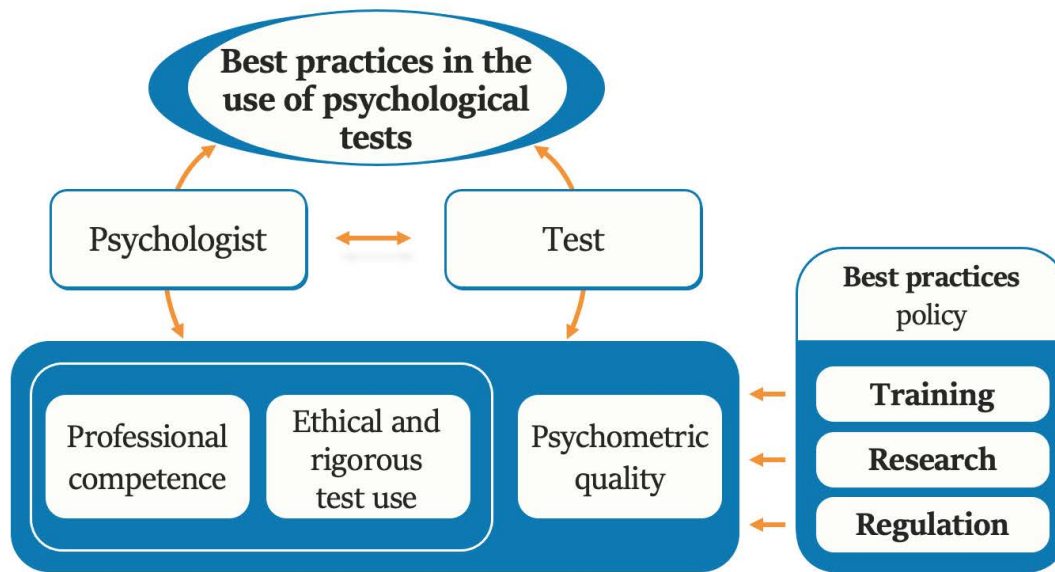
Type	Source (author, year, title)	Summary of its contribution
Technical Guidelines	ITC. (2013). International Test Commission Guidelines on Test Use.	The International Test Commission developed a set of guidelines on the good use of tests that work on psychological, educational, and occupational measurement, promoting fair, valid, transparent, and efficient practices, guidelines, and policies.
	APA. (2020). APA Guidelines for Psychological Assessment and Evaluation.	A set of guidelines on psychological assessment from a scientific and professional organization representing psychology in the United States that promotes the advancement, communication, and application of psychological science and knowledge.
Manual	AERA et al. (2018). Estándares para pruebas educativas y psicológicas (M. Lieve, Trad.). American Educational Research Association.	A manual published collaboratively by the three organizations since 1966, it represents the international “gold standard” in testing guidelines.
Books	Fernández-Ballesteros et al. (2011). Buenas prácticas y competencias en evaluación psicológica. El Sistema Interactivo Multimedia de Aprendizaje del Proceso de Evaluación (SIMAPE).	This book describes a multimedia system designed for training in the technical competencies needed to conduct an ethical psychological assessment process based on the scientific method.
	Vinet & Rodríguez-Cancino. (2024). Evaluación Psicológica en Chile: Historia, Actualidad y Desafíos.	The book chapter discusses the history of psychological assessment in Chile, the difficulties observed, and the need to improve its practice.
Scientific articles	Elosua. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en psicología.	An article that discusses projects related to test use and the importance of committees and guidelines to improve their use.
	Evers et al. (2012). Testing Practices in the 21st Century: Developments and European Psychologists’ Opinions.	An article summarizing actions and projects of the European Federation of Psychologists’ Associations (EPPA) to improve assessment practices in European countries.
	Fernández-Ballesteros et al. (2003). Guías para el proceso de evaluación (GAP): Una propuesta a discusión.	An article describing the development of the Guidelines for the Assessment Process (GAP), sponsored by the European Association for Psychological Assessment (EAPA) as a proposal on the need to improve the practice of psychological assessment and training in the area.
	Hernández et al. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test.	This article reviews the main activities and projects carried out by the Test Commission of the General Council of Psychology of Spain to improve the quality and use of assessment instruments.
	Muñiz et al. (2011). Evaluación de test editados en España.	The article reports the first test evaluations using an assessment model developed by the European Test Commission in the Spanish context.
	Muñiz et al. (2020). El uso de los Test en España y Latinoamérica: Perspectivas actuales y retos futuros.	The article provides an updated overview of testing practices in Spain and Latin America and their future challenges.
	Prieto & Muñiz. (2000). Un modelo para evaluar la calidad de los tests utilizados en España.	Article describing a model of assessment and dissemination of the technical aspects of tests as a means to improve their use by the Colegio Oficial de Psicólogos de España (Official Association of Psychologists of Spain).
	Vinet et al. (2023). El Empleo de Test por Psicólogos/as Chilenos/as: Un Inquietante Panorama.	This article presents the results of the only study (to date) in Chile that has addressed the use of tests, the evidence of their psychometric quality, and proposals for improving best practices.
	Wechsler et al. (2019). O desenvolvimento da avaliação psicológica no Brasil: Avanços históricos e desafios.	This article addresses the history of psychological assessment in Brazil, identifying the foundation of the Brazilian Institute of Psychological Assessment and the System of Evaluation of Psychological Tests (SATEPSI) for improving the use of tests as relevant milestones.

## 4. Results

The development of the topic presented in this review is structured according to the three essential pillars mentioned above: professional competence and the ethical and rigorous test use, which focuses on who is administering the instrument, and the psychometric quality of the tests, which emphasizes the instrument itself. For each of these central concepts, considerations are included for a policy of best practices in using tests in the Chilean context, which are developed in the section on recommendations for action. Figure 1 illustrates and summarizes the approach to the subject.

**Figure 1**

Diagram of best practices in the use of psychological tests



Source. Prepared by the author

#### 4.1 Professional competency: Solid and relevant training in the use of tests

The rapid and continuous development of instruments, procedures, regulatory advances, technology, and the evolution of evidence-based practices obliges psychologists to seek ongoing training in their assessment procedures (APA, 2020). Tests are constantly being updated, and empirical evidence generates data that should be incorporated into their use; therefore, the psychologist must be educated, trained, and supervised by peers or those with more experience using a specific test.

Competency in using tests refers to a solid knowledge of their theoretical and empirical principles. This means that the psychologist must be an expert in the proper procedures for administering and correcting the tests they work with, as well as in the theoretical, empirical, and psychometric foundations that support the interpretation of their scores.

According to Forns and Amador (2017), the proper administration of tests and assessment instruments requires that the psychologist be competent in:

- (1) cognitive skills: knowledge of assessment instruments, their theoretical basis, constructs to be assessed, and areas of application;
- (2) procedural skills: knowledge of the standards and rules of administration, scoring, interpretation of results, and specific training in the tests that require it; and
- (3) clinical skills: skills in observing the verbal and nonverbal language of the person being assessed during the administration of the instruments and the ability to relate their behavior to the results, the theoretical foundations of the tests, and the possible presence of changes or dysfunctions.



On the other hand, APA guidelines (2020) indicate that psychologists should strive to develop and maintain skills to select, use, and interpret tests, integrate findings with other data sources relevant to the reason for the consultation, and properly communicate the results.

Psychologists must ensure their proficiency in test administration and only provide services in those for which they have the requisite training. Additionally, psychologists must know the limits of their own competency and not act beyond them. This implies the obligation to keep abreast of changes and advances in the construction and use of tests and to be fully aware of the legal frameworks governing their use (Fernández-Ballesteros et al., 2011; ITC, 2013).

Some tests demand considerable effort from the evaluators, such as the rapid recording of answers, rigorous and standardized handling of materials or equipment, or the implementation of items and delivery of highly complex instructions. In this light, the guidelines of the AERA et al. (2018) state that, if the examiners do not have and/or cannot develop the necessary skills to comply with these requirements correctly, they should be aware of their personal limitations and not administer such tests.

#### ***4.2 Quality of tests: assessment of their psychometric properties***

Professionals who use educational or psychological tests and must make relevant decisions with the information derived from them must ensure that their evaluations are scientifically substantiated to support their conclusions and recommendations (Sireci & Benítez, 2023). In this area, having information available regarding the psychometric properties of tests (validity evidence, reliability, and fairness) and verifying their quality through them can safeguard their use as appropriate and fair (AERA et al., 2018; ITC, 2013; Muñiz & Fonseca-Pedrero, 2019; Vinet et al., 2023).

##### **4.2.1 Validity evidence**

The psychometric concept of validity refers to the degree to which evidence and theory support the intended interpretation of a test's scores for the intended use (AERA et al., 2018). This definition assumes that validity is not an inherent property of the test and it refers to the proposed use and interpretation of the scores. Consequently, scores may be valid for one purpose (supported by empirical and theoretical evidence) and invalid for another (Elosua, 2012; Leong et al., 2020; Sireci & Benítez, 2023).

Acquiring validity evidence of a test consists of collecting and analyzing evidence to defend its use for a given purpose and interpreting its scores (Elosua, 2017; Sireci, 2020). Leong et al. (2020) point out that validation is a process of building and evaluating evidence for or against the interpretation of test scores for a given purpose. This process, which begins before the test's construction and guides its development and validity, includes a set of studies that make a coherent theoretical interpretation of its scores possible (Elosua, 2003; Muñiz & Fonseca-Pedrero, 2019). The review of the evidence collected in validation processes is especially relevant for a psychologist when selecting a test for a specific purpose (Vinet et al., 2023).

It should be noted that according to the guidelines of the AERA et al. (2018), validation is a shared responsibility between the test developer and the professional using it. On the one hand, the developer of the instrument must collect and provide relevant evidence to support any interpretation of its scores within the framework of the intended purpose. On the other, the psychologist who decides to use it must be able to assess the relevance of this evidence in the particular context in which the test will be used.



Several authors (Brenlla et al., 2023; Elosua, 2003; Iliescu et al., 2024; Leong et al., 2020; Sireci & Benítez, 2023) agree with the classification proposed by AERA et al. (2018), who point out that the sources of validity evidence derive from the content of the test, its internal structure, its relation to other variables, the response process or the consequences of its application.

Table 2 describes in simple terms some examples of the scope and nature of each source of validity evidence.

**Table 2**

*Summary of sources of validity evidence considered in the current definition of validity*

Sources of evidence of validity based on:	Makes verification possible if:	Examples
Content	The items of a test <b>adequately represent the diversity of aspects</b> of the construct to be measured.	A test of knowledge of the four arithmetic operations; the items probe addition, subtraction, multiplication, and division (high content validity) versus another test where the items probe only addition and subtraction (low content validity).
Internal structure	The items and subscales of a test confirm <b>the theoretically expected pattern of internal associations</b> .	Empathy test items are empirically organized into two dimensions theoretically identifiable as cognitive empathy and affective empathy (high structural validity) versus being organized into a single dimension (low structural validity).
Relations to other variables	Test scores are <b>associated in the theoretically expected manner</b> with (a) other measures of the same construct (concurrent validity), (b) measures of other constructs (convergent validity), or (c) are not associated with measures of other constructs (discriminant validity).	Scores on a depression test show a positive correlation with another depression test (high concurrent validity), a positive correlation with an anxiety test, a negative correlation with an optimism test (convergent validity), or no correlation with a social desirability test (discriminant validity).
Response process	Responses to the test items <b>actually activate the psychological processes</b> that should theoretically be activated.	A test that measures family support assumes that all respondents share the same concept of family (people with kinship relations). If people living alone include acquaintances or neighbors in their family, then this test has a low response process validity.
Consequences of test use	The interpretation or use of test results <b>has intended and unintended consequences</b> .	An assessment of pedagogical excellence generates in favored teachers a higher professional self-esteem (intended consequence: high validity) but little recognition from their peers, managers, and sponsors (unintended consequence: low validity).

#### 4.2.2 Measurement reliability/precision

Any professional who uses tests as one of the sources of information that will guide decision-making must ensure that their scores are reasonably reliable, i.e., consistent and accurate (Urbina, 2014). Reliability is a fundamental condition of a measurement instrument and an indicator of its quality. If a test has low reliability, its scores will be unstable or inconsistent, primarily reflecting measurement error (Mena, 2019; Iliescu et al., 2024).

Reliability is not a question of “all or nothing” but rather one of degree, and it is more advisable to see it as a continuum that can go from the minimum consistency of a measurement to the almost perfect replication of the results (Gregory, 2012). Professionals who use tests should examine this property and select only those with at least an acceptable level of reliability.

The assessment of a test's reliability is a crucial factor that ensures its scores are helpful, meaning they are sufficiently consistent and devoid of measurement errors (Urbina, 2014), assuring the professional that their diagnostic or intervention decisions have a robust psychometric foundation. It is important to consider that reliability/precision may vary from one population to another, as sources of error such as rater effects, item familiarity, or test materials may have a greater impact on one population than another (AERA et al., 2018). Therefore, it must not be assumed that if the psychometric exploration of a test yields good reliability indices for a particular sample of people, they will maintain their quality if applied to others who differ in their personal or cultural characteristics. This must be examined and checked.

The reliability/precision of a measurement depends on several factors that have to do with who performs the measurement (examiner), what is used to perform it (measurement instrument-test), and what is being measured or observed (Manterola et al., 2018). Gregory (2012) points out that scores on a test will always be the result of factors that contribute to consistency (stable attributes intended to be measured) and factors that contribute to inconsistency (characteristics of the person being assessed, the test, or the assessment situation), which are not related to the attribute to be measured, but which affect the scores. According to this author, this means that the score obtained by a person assessed on a test (observed score) always incorporates the measurement of the attribute, plus the elements of measurement error, so that the true score (without errors) can never be known. Therefore, test scores only represent an estimate of the true score of the attribute being measured, which will tend to be better as the measurement increases in reliability.

Measurement error in test usage refers to any variation in test scores caused by elements inherent in the assessment process unrelated to the measured attribute. When reviewing the psychometric properties of tests and their degree of reliability, it is important to remember that these sources of measurement error reduce the usefulness of a test's scores and the possibility of relying on the precision of the results; therefore, reliability and measurement error are two interdependent aspects closely linked to the consistency of a measurement instrument (AERA et al., 2018; Gregory, 2012; Urbina, 2014).

#### 4.2.3 Fairness

This psychometric property can be understood as the sensitivity to respectfully consider all the characteristics of the people assessed (e.g., ethnicity, gender, age, socioeconomic level, linguistic, or cultural context) during all stages of development, application, scoring, and interpretation of psychological tests (Díaz-López et al., 2017). A particular test is not necessarily able to measure the same constructs in people belonging to different subgroups since the content or format of the instrument may differentially affect test takers' scores, favoring or disfavoring them unintentionally. Therefore, if a decision is made to use a test in a particular subgroup, the existing instrument validation evidence for that subgroup should be rigorously analyzed (AERA et al., 2018; Brenlla et al., 2023).

A test is fair if it reflects the same constructs for everyone assessed without favoring or disadvantaging anyone due to personal characteristics or affiliation with a specific group in the population. When this does not occur, a central threat to fairness emerges: the measurement bias (AERA et al., 2018). Regarding the issue of managing possible biases in tests, the ITC guidelines (2013) establish that if a test is used with people from different groups (according to gender, culture, educational level, ethnicity, etc.), it is the psychologist's responsibility to ensure that they are fair and appropriate for them. To do this, the professional must ensure that the constructs being measured are relevant to each of the groups and must be able to collect and review evidence on the differential functioning of the items and/or validation processes that support their use in different groups.

These guidelines also state that when tests are used in more than one language, it is necessary to ensure that all content, cultural, and idiomatic aspects have been thoroughly considered in the construction of the different versions and that the person administering the test can communicate perfectly in the language in which the test is to be administered.

### **4.3 Ethical and Rigorous Use of Tests**

The proper use of tests implies that the psychologist can select, administer, correct, and interpret them correctly and responsibly. This encompasses the professional's obligation to ensure their care and safeguard their integrity by upholding their copyright (prohibition on reproduction), preventing the dissemination or publication of their materials in mass media or social networks, and ensuring that the instruments are utilized solely by individuals possessing the requisite training and authorization (Hall et al., 2005; ITC, 2013). These aspects, which refer to the security of the tests, are fundamental to guaranteeing the validity of the interpretations of their scores. For example, there has been an increase in the availability of technologies for photographing or sharing records of test items or attempts to obtain tests before answering them. This puts the potential usefulness of the measurement at risk, as it will not correctly reflect the construct it intends to measure, which can have serious consequences for the people being assessed and for decision-making (Muñiz et al., 2015).

#### **4.3.1 Test selection**

Tests must be chosen based on the rationale for the consultation, the goals of the assessment, and the hypotheses formulated throughout the evaluation process (Forns & Amador, 2017). This task requires the use of assessment tools with sound psychometric quality, specifically those that possess validity evidence, acceptable score reliability, and the ability to produce measures that are fair and suitable for the assessment's purpose, the population, the setting, and the context in question (APA, 2020).

According to the ITC (2013) guidelines, psychologists must evaluate the potential usefulness of tests for each assessment situation, reasonably justifying their use and only after performing a rigorous analysis of the needs and characteristics of the person to be assessed and always considering additional sources of information. The professional should weigh the advantages and disadvantages of choosing a particular test against other information sources.

The ITC (2013) indicates that the psychologist must choose technically adequate tests relevant to each assessment situation. To do this, they should examine all available information and check that the technical documentation on the test includes information on the representativity of its content, characteristics of the normative groups used in the scoring, reliability/precision of the measure, validity evidence, and absence of biases supporting its use. The professional should not use instruments that have inadequate or unclear technical documentation and should not accept a test based solely on its face validity, i.e., the illusion that something is useful or relevant just because it is widely used, recommended by other users, or on the advice of those with commercial interests.

In test selection, it is essential to consider the validity and reliability/precision evidence described above, as well as the demographic characteristics of the groups for which the test was originally designed. Selecting tests with demographic and clinically relevant standards for the person being assessed and for the objective of the assessment makes it possible to support the inferences made after their application. It is not appropriate to apply a test designed for a specific group to other groups without first examining its relevance (AERA et al., 2018).

#### **4.3.2 Test administration**

A correct administration of psychological tests requires that the evaluator develop cognitive, procedural, and clinical skills (Forns & Amador, 2017), in addition to ensuring adequate internal and external conditions for the person being assessed.

#### *4.3.2.1 Rapport*

According to the ITC (2013) guidelines, appropriately applying the tests first requires that the evaluator establish a good relationship with the persons being assessed, address them positively, and try to reduce their anxiety about the assessment situation. Furthermore, they must be mindful of removing any distractions, including alarm clocks or cell phones, and ensure that all materials necessary for the test administration are prepared in advance.

Fostering a positive relationship with the individual being assessed calls for establishing a rapport, which Forns and Amador (2017) define as listening actively, adapting to the specific characteristics of the person being assessed, and respecting their timing, among other factors. Establishing rapport is a crucial skill for examiners, as the quality of rapport can influence individuals' performance on standardized tests, impacting professional decisions and significantly affecting people's lives (Hall et al., 2005).

#### *4.3.2.2 Performance Observation*

Another element of great relevance in test administration is the evaluator's ability to observe and record. Observing and describing the verbal and nonverbal behavior of the person being assessed, their attitude towards the assessment situation, problem-solving strategies, affective disposition, attention level, mood, or sensory and motor skills are an indispensable complement to the interpretation of the scores that emerge from a test (Cortés & Benavente, 2007; Echavarría-Ramírez & Tirapu-Ustárrroz, 2021; Rodríguez-Cancino et al., 2024; Sattler, 2010).

Experts in psychological assessment processes have stressed the importance of looking beyond test scores. One of the main advantages of individually administered psychological tests is the ability to closely observe an individual's performance, allowing the examiner to determine whether the scores accurately reflect a person's ability on a standardized task. Developing this skill requires solid training (Hall et al., 2005; Rodríguez-Cancino et al., 2024).

#### *4.3.2.3 Environmental conditions*

Regarding external conditions for proper test administration, the guidelines of the AERA et al. (2018) point out that test developers must provide all the information about the environmental conditions required to apply an instrument, while it is the evaluator's responsibility to ensure compliance. In addition, they explicitly point out that an adequate measurement of a construct cannot be achieved if the working space is inadequate, noisy, has unsuitable temperature and light, or low-quality or illegible materials are used.

#### *4.3.2.4 Standardized procedures*

Another important aspect in properly administering tests is the rigorous compliance with standardized procedures. Standardized tests use the same materials, administration, and correction rules for each person tested. The significance of strictly following these rules and refraining from modifications is rooted in their role as a mechanism that promotes fairness, facilitates score comparisons between individuals, and guarantees the use of normative data. In tests designed to assess knowledge, skills, abilities, or other personal characteristics, standardized procedures ensure that all test takers have an equal opportunity to demonstrate their competencies, and no one has an undue advantage (AERA et al., 2018).

Non-compliance with or deviation from standardized procedures, in theory, decreases the reliability and validity of the test results since the accuracy of the measurement obtained once the test is applied not only lies in the instrument but also in the circumstances of its use (Hall et al., 2005). If there are occasions when it becomes necessary to adjust or modify standardized test administration procedures, the professional should be aware that any modification will bring about changes in the estimate of the construct the instrument originally measured. Hence, any modification that changes the construct will immediately invalidate the possibility of using standards to interpret those scores (AERA et al., 2018). In such instances, prior to implementing a modified test, it is essential to collect any data required to assess the validity of the interpretations for the anticipated applications of the scores.

#### **4.3.3 Interpretation and Communication of Test Results**

After the appropriate selection, administration, and correction of the test, the professional enters into the complex task of interpreting the results, synthesizing all relevant data sources, and then conveying the results. This process is highly relevant to the person being assessed since this information will be used to make referral, intervention, and/or treatment planning decisions. In this task, the psychologist must integrate, globally and comprehensively, all the data collected in the assessment process together with their knowledge of psychology, psychopathology, and psychometrics, and employ appropriate linguistic and communicative skills (Cortés & Benavente, 2007; Fernández-Ballesteros et al., 2011; Forns & Amador, 2017; Rodríguez-Cancino et al., 2024).

According to the guidelines of the AERA et al. (2018), test manuals indicate the variables that should be considered when interpreting scores, highlighting the person's clinically relevant history, medication use, school, work, or vocational history, among others. In addition, they emphasize that background related to age, culture, disability, gender, and linguistic or racial/ethnic characteristics should always be considered in a relevant manner.

For its part, the ITC (2013) guidelines indicate that for psychologists to interpret the results of a test adequately, they must have a deep understanding of its theoretical/conceptual foundations, the types of scores used, and the standards, as well as the limitations of the measurement instruments they have administered. In the interpretation task, the professional must try to minimize any possibility of bias and consider all available information about the person being assessed (e.g., age, gender, schooling, or culture), avoiding the perpetuation of stereotypes or making generalizations of the results to traits or characteristics of the person that the test has not measured. It is also underscored that in the interpretation of the scores, the validity evidence, reliability, measurement error, and all those elements that may artificially alter the results must be considered.

When making reports (oral or written) of test score interpretations, emphasis should be placed on appropriately using this information and minimizing the potential negative consequences. The report of the accuracy of the scores should be accompanied by the margins of error or probable ranges and include references to all factors that could affect the results and interpretations. The psychologist is responsible for avoiding misinterpretation and ensuring that results or scores are not misused (AERA et al., 2018).

Regarding the communication of the results, the ITC (2013) indicates that the psychologist must be able to do so clearly and precisely, either orally or in writing, adjusted to the recipient's characteristics, interests, and roles. It is also indicated that the professional must be able to safeguard the confidentiality of the data and identify the persons or institutions with whom the results of an evaluation process are shared, always with the explicit consent of the person being assessed.

## 5. Recommendations for action

### *5.1 Considerations for training in test use in Chile*

According to this background and the current scenario illustrated by Vinet and Rodríguez-Cancino (2024), a policy that promotes best practices in the use of tests in Chile should consider strengthening undergraduate training and the need to create graduate programs that enable professionals to cultivate these professional competencies adequately.

The strengthening of initial training requires that universities that offer psychology programs agree on minimum standards and content based on international guidelines and national training needs. In this regard, it is necessary and mandatory to at least include adequate training in psychometrics to equip professionals with the ability to select the most relevant tests for specific consultation purposes; training in clinical skills to administer the instruments correctly; and consolidation of a robust conceptual framework (for example, in affective and cognitive processes, human development, normative and non-normative trajectories, psychopathology, etc.), which allows psychologists to interpret and integrate the results from various measurement instruments effectively.

On the other hand, in line with Juliá's (2023) proposals, initial professional training should mandatorily promote the need and value of constant ethical reflection on professional practices, advocating for the integration of evidence-based practice with permanent reflection on the decisions made and actions undertaken by psychologists.

At the graduate level, and given the obligation of permanent updating, it would be necessary to have instances of specialization in areas of assessment or specific tests. The advancement in psychometrics and the oversight of psychological evaluation procedures where tests are used exemplify this, provided they incorporate mechanisms that guarantee the quality of the training offered. In this sense, all those instances associated with universities with quality assurance mechanisms, whether in graduate training or continuing education, would be the most reliable and suitable setting to develop this proposal.

Training in the proper use of tests at the undergraduate and graduate levels should also reinforce the psychologist's ability to assess the psychometric quality of the measuring instruments they choose to use in a psychological assessment process. This means at least developing the ability to understand the evidence of validity, reliability, and fairness and interpret it appropriately. When selecting a test, only in this way can an informed decision be made regarding whether it is useful as a measuring instrument for a particular assessment.

### *5.2 Considerations for test quality research and study in Chile*

The psychometric properties of tests certify their quality as suitable measurement tools; they must be constantly evaluated to ensure their use in a particular population. As mentioned above, the research and development of instruments have been gradually strengthened in Chile; however, it is necessary to have national policies that promote their consolidation, so they do not stagnate (Vinet et al., 2023).

The study by Vinet et al. (2023) found that Chilean psychologists widely use projective tests that lack evidence regarding their psychometric properties and/or cultural relevance. They also found that of the 20 most commonly used tests nationally, only five

had this type of evidence. Therefore, psychometric research in Chile should be oriented to constantly evaluating the quality of all the tests in widespread use in terms of their accuracy, relevance, and fairness, guaranteeing their usefulness for the intended applications.

Consequently, it is imperative that accredited institutions and organizations (such as universities, research centers, or professional societies), in continuous collaboration with entities involved in professional practice (e.g., the Association of Psychologists), ensure the establishment or preservation of specialized research teams dedicated to generating and disseminating evidence regarding the psychometric properties (validity, reliability, and fairness) of commonly utilized tests in professional practice, to either endorse or restrict their widespread application.

In addition to promoting research on the psychometric properties of the tests, this task should include mechanisms that can assess their quality, guaranteeing that their use is fair and relevant for the Chilean population, especially for those instruments on which decisions are based that have a profound impact on people's lives, whether in the clinical (diagnosis or treatment), educational (admission to specific support programs), occupational (selection of people) or judicial (counseling or expert opinions) spheres. In this sense, it would be advisable to emulate the approach taken by countries such as Spain (Hernández et al., 2022; Prieto & Muñoz, 2000) or Brazil (Wechsler et al., 2019; Wechsler et al., 2022) which have implemented test evaluation systems that certify their technical-scientific quality based on adherence to minimum theoretical and psychometric requirements, thereby facilitating or limiting their professional application.

It is important to note that those in charge of editing and/or marketing the tests also play a fundamental role. A policy of best practices regarding the use of instruments in Chile should incorporate strategies that ensure the evaluation and oversight of whether these entities consider the assessment of the psychometric quality of the instruments and have mechanisms to guarantee their relevance to the local context.

Finally, it is essential that undergraduate and graduate training be aligned with psychometric research findings. Given the widespread use of projective tests without evidence of their relevance to the Chilean population, it is reasonable to assume that psychologists are being trained in their use. Therefore, it would be advisable for training plans and programs to rigorously select the tests to be taught, favoring those that have a psychometric quality assessment sufficiently adequate to guarantee their usefulness.

Despite the acceptable demonstration of a test's psychometric qualities, its effectiveness as a measurement instrument may diminish if administered by an individual lacking the requisite competencies (Muñoz et al., 2011). It can also lose its effectiveness if the sources of measurement error are not controlled (as far as possible) or if the psychologist does not use it ethically and rigorously.

### ***5.3 Considerations for the regulation of the ethical and rigorous use of tests in Chile***

International evidence has shown that the ethical and rigorous use of tests requires mechanisms to regulate professional practice. However, as Vinet and Rodríguez-Cancino (2024) point out, these mechanisms are currently nonexistent and highly needed in Chile. In this regard, it should be noted that although it is true that today, the Association of Psychologists cannot exercise control over professional practice, progress is being made at the national level to change this through a bill that aims to return this power to all professional societies (Carrillo, 2024). Should this initiative yield a positive result, implementing measures such as compulsory collegiality, accreditation of training in specific tests, and sanctions for ethics violations would be expected.



It should be noted that the ethical and rigorous use of tests cannot rely solely on external regulatory mechanisms. Psychologists must also self-regulate their practices based on a deeply developed personal and professional ethical framework. This aspect must be a priority in the training of initial professional competencies.

It would also contribute to the ethical and rigorous use of tests to examine the validity evidence based on the consequences of their application. This process would make it possible to determine the impact or implications for the people assessed and could draw attention to bad practices and inadequate use of the results, among others.

The complement of internal and external regulatory mechanisms would aid in making significant progress in ensuring that psychologists comply with minimum standards for the adequate selection, administration, and correction of tests, as well as interpretation and communication of their results. Creating a best practices policy for the use of tests would facilitate the effective alignment of all the current national challenges in training, research, and the regulation of professional practice, as previously outlined.

## 6. Limitations and future directions

One of the limitations of this work is that, due to the nature of a narrative review, the selection of data sources is often subjective, influenced by the authors' perspectives or interests, which may introduce biases. It would be advisable to review the practices on using tests in Chile and/or Latin America through a systematic review or meta-analysis to counteract possible sources of bias. It is also recommended that the empirical approach to this issue be continued following the framework established by the Vinet et al. team (2023) to identify new needs and challenges. Accelerated social and technological changes are strongly impacting psychometrics with the incorporation of artificial intelligence, virtual reality, and the need to computerize measurement instruments. This scenario necessitates a reevaluation of current practices in the use of tests and reflection on the regulatory mechanisms to be implemented.

## 7. Conclusions

Best practices in the use of tests minimally require that the psychologist have solid training in the area, that the quality of the instruments be guaranteed, and that they be used ethically and rigorously. This paper addressed a brief review of the main international guidelines related to the three pillars identified by several authors as essential aspects to ensure the good use of psychological tests (Elosua, 2017; Evers et al., 2012; Hernández et al., 2022; Muñiz et al., 2011; Muñiz et al., 2020) and the elements that should be considered in the design and implementation of a policy that promotes best practices regarding test use in Chile.

First, the importance of test administration by competent and qualified professionals regarding their cognitive, procedural, and clinical skills was highlighted, emphasizing the need to strengthen undergraduate and graduate training in this area. On the one hand, this aspect mainly invites universities and other institutions to coordinate, establish agreements, and verify their compliance in offering specialized and quality training in the area. On the other, it underscores the importance of psychologists constantly evaluating their possibilities and limitations to strengthen their competencies and permanently update their knowledge.

Second, some psychometric properties (validity and reliability evidence, and fairness) that should be guaranteed and that indicate the quality of the tests as reliable measurement instruments were described. This task is mainly the responsibility of research teams in the area, whether from universities, laboratories, research centers, or scientific societies. However, this is also the responsibility of the psychologist who uses the tests, since when selecting the instruments to be used in a psychological assessment process,

they must be able to collect and evaluate the psychometric evidence available and thus decide whether they are useful to support their inferences.

Finally, the elements that minimally represent the ethical and rigorous use of tests were developed, alluding to the essential aspects that should be considered in the appropriate selection and administration of instruments, as well as the interpretation and communication of their results. These aspects are based on the first two points reviewed and fundamentally rely on the ethical conduct of the psychologist while also necessitating an environment conducive to establishing suitable regulatory systems for professional practice.

An example of the interconnection among the topics discussed in this work can be seen in the approaches of Juliá (2023) in educational psychology. The author emphasizes the need to incorporate consideration of the context to design interventions that are culturally relevant and grounded in this area of professional practice. Within this, she points out that psychological assessment processes in education contexts should respect diversity, guarantee equity, and ensure they do not cause discrimination or segregation. The author asserts that it is an ethical imperative to consider the examinees' rights, as outlined in international standards, throughout the evaluation process, as this forms the basis for a "chain of education and family decisions" that affect the educational trajectory of children and adolescents. A national policy that addresses and regulates the use of tests in the psychological assessment process would contribute to meeting the author's requirements ethically, responsibly, and rigorously.

It is hoped that the ideas developed in this text will illuminate and encourage the initiation of a policy for best practices in test use in Chile. This work requires the collaboration and ethical commitment of various individuals and institutions, which will undoubtedly improve the professional practice of psychology while simultaneously safeguarding the right of all people to a fair and relevant assessment.

# ***Consideraciones para una política de buenas prácticas en el uso de pruebas psicológicas en Chile***

## **1. Introducción**

El término “evaluación” se refiere a la compleja actividad de integrar el conocimiento, juicio clínico, diversas fuentes de información y constructos psicométricos en un proceso de resolución de problemas (American Psychological Association [APA], 2020). La evaluación psicológica es un proceso con etapas o fases específicamente definidas, que permiten orientar la toma de decisiones a través de un conjunto de acciones científicas y profesionales para la recolección, valoración e integración de información sobre una persona. Este proceso se lleva a cabo a partir de una planificación de acciones en las que es de alta importancia la utilización de diversas fuentes de información (Fernández-Ballesteros et al., 2003; Fernández-Ballesteros et al., 2011; Hernández et al., 2021; Marín, 2021).

Dentro de estas fuentes de información se encuentran los test, que se pueden definir como cualquier dispositivo o instrumento de medición que, a través de las respuestas de una persona a una serie de ítems, pretende evaluar constructos psicológicos. Los test buscan capturar una muestra estructurada del comportamiento de una persona en un dominio específico, con el objetivo de orientar la toma de decisiones en diversos campos de la psicología (APA, 2020; Ferrando et al., 2022; Muñiz et al., 2020). Los test pueden ser instrumentos centrales dentro de un proceso de evaluación psicológica, sin embargo, para que puedan configurarse en herramientas fiables de apoyo al quehacer profesional deben ser utilizados adecuadamente (Prieto & Muñiz, 2000).

El uso adecuado de test psicológicos está resguardado por entidades internacionales tales como la American Education Research Association (AERA), la International Test Commission (ITC), y la APA, quienes cuentan con un amplio reconocimiento de su contribución científica en el área. Estas organizaciones se han encargado de consensuar, generar e implementar medidas y directrices en este ámbito (Díaz-López et al., 2017; Hernández et al., 2021).

### **1.1 Bases del adecuado uso de test y contexto chileno**

Diversos autores coinciden en que la utilidad de los test se asienta en la adecuada articulación de tres pilares esenciales (Elosua, 2017; Evers et al., 2012; Hernández et al., 2021; Hernández et al., 2022; Muñiz et al., 2011; Muñiz et al., 2020):

- 1) la administración de los test debe estar a cargo de profesionales competentes y calificados/as,
- 2) la calidad (propiedades psicométricas) de los test debe estar garantizada, y
- 3) el uso de los test se debe regir por la ética y la rigurosidad.

El primero de estos pilares alude esencialmente al campo de la formación profesional. El segundo, es una tarea que principalmente desarrollan los equipos de investigación en el área. El tercero, se refiere a un ámbito que se debiese cautelar a través de adecuados mecanismos de regulación del ejercicio profesional.

El escenario actual en Chile evidencia grandes desafíos en torno a estos tres pilares. En relación a la formación profesional, Vinet y Rodríguez-Cancino (2024) señalan que en las últimas décadas ha habido una alta proliferación de programas que ofrecen la

carrera de psicología a nivel de pregrado, que cuentan con escasa formalización y/o sin mecanismos mínimos de aseguramiento de calidad. Las autoras mencionan que han existido esfuerzos para asegurar la calidad de la formación, tales como la creación de la Red de Escuelas de Psicología del Consorcio de Universidades Estatales de Chile, la cual ha establecido que la evaluación psicológica es una de las macro-competencias centrales para pregrado. Sin embargo, por una parte aún no existe un total acuerdo ni regulación en torno a los contenidos de evaluación psicológica y psicometría que se debiesen impartir. Por otra parte, en el estudio de Vinet et al. (2023) se reporta que los/las psicólogos/as encuestados/as consideran que la formación inicial no ha sido suficiente para el adecuado uso de test en el mundo profesional. De acuerdo con esto, en Chile se aprecia la necesidad de fortalecer la formación en esta área, otorgándole un carácter especializado y/o de postítulo y posgrado.

En cuanto al segundo pilar, Vinet y González (2013) mencionan que los déficits a nivel de formación en Chile mermaron la creación o adaptación de test, lo que se ha ido compensando gracias al aumento de postgraduados/as. Según Vinet et al. (2023) la investigación y desarrollo de instrumentos en Chile se ha visto enriquecida tras la creación de instituciones nacionales que se encargan de estas tareas, sin embargo, aún es urgente fortalecer equipos de investigación que se dediquen a la construcción, traducción y/o adaptación de instrumentos, aplicando rigurosamente las directrices internacionales. Además, señalan que es necesario instalar mecanismos que puedan valorar la calidad psicométrica, las normas y la pertinencia cultural de los test.

Por último, el tercer pilar alude la necesidad de contar con mecanismos de regulación de la práctica profesional. Es importante considerar que en Chile en la década de los 80 el Colegio de Psicólogas y Psicólogos sufrió la pérdida de sus funciones de tuición legal sobre el ejercicio profesional generando un escenario de praxis sin regulación (Vinet et al., 2023; Vinet & González, 2013). Esta situación ha facilitado que en el país exista un gran uso de instrumentos que no cuentan con respaldo en torno a sus propiedades psicométricas ni evidencia sobre la pertinencia para la población chilena, por lo que se hace indispensable contar con estrategias que regulen las prácticas tales como sistemas acreditación en el uso de test, directrices sobre qué test se pueden utilizar y cuáles no, y sanciones ante el incumplimiento de condiciones mínimas para su uso (Vinet & Rodríguez-Cancino, 2024; Wechsler et al., 2022).

Estos desafíos a nivel de la formación, investigación y regulación sobre el adecuado uso de test fundamentan la necesidad de comenzar a trabajar en medidas que garanticen buenas prácticas dentro de los procesos de evaluación psicológica en Chile.

### ***1.2 ¿Por qué es necesaria una política sobre buenas prácticas en el uso de test?***

Las personas evaluadas tienen derecho a que se les apliquen instrumentos que cumplan con los estándares profesionales de calidad técnica, en condiciones apropiadas de administración, prevaleciendo la objetividad e imparcialidad en la interpretación y reporte de los resultados obtenidos. La inadecuada utilización de test en procesos de evaluación psicológica atenta principalmente ante estos derechos y tiene serias consecuencias para las personas evaluadas, dado que a partir de ellos se toman decisiones relevantes para sus vidas (AERA et al., 2018; Muñiz & Hambleton, 1996).

Las investigaciones que han abordado la percepción de problemas en torno al uso de test coinciden en identificar, entre los más comunes, la utilización de éstos por parte de personas no cualificadas, uso de instrumentos inadecuados, desconocimiento de la calidad psicométrica, o fotocopiar materiales sujetos a copyright. También se ha detectado el uso de test foráneos con procesos de adaptación deficientes o ausencia de estudios de validación (Evers et al., 2012; Hernández et al., 2021; Muñiz et al., 2020; Prieto et al., 1999).

En Chile, el estudio de Cataldo et al. (2012) exploró la calidad psicométrica del Test de Alerta, instrumento que se utiliza ampliamente en los procesos de selección de personas, para evaluar la propensión al riesgo en el ámbito de la minería. Dentro de su reporte identifica diversos problemas en el uso, tales como que no cuenta con una definición de los constructos psicológicos que pretende evaluar, ni manual de administración y que se utiliza principalmente a través de fotocopias sin respetar los derechos de autor. Además, señala que este test posee ítems descontextualizados al sector en que se aplica y que carece de normas de interpretación de los resultados, siendo cada organización y/o profesional que lo utiliza quienes establecen, de manera arbitraria, un puntaje para selección. Los resultados de esta investigación demostraron que el test no cuenta con evidencia sobre su nivel de confiabilidad o validez, y los/as autores/as sugieren discontinuar su uso dada su inexactitud y el riesgo de utilizarlo para estimar la posibilidad de accidentabilidad de las personas. Además sugieren que los/as psicólogos/as no deberían mantener una actitud pasiva en cuanto a utilizar test de los cuales se desconoce su origen y calidad como instrumento de evaluación.

Otro estudio a nivel nacional evidenció un amplio uso de test y técnicas para la selección de personas que no cuentan con una base científica que respalde su utilización y capacidad predictiva. Cabe destacar que este estudio detectó que estos test son frecuentemente requeridos desde las empresas, solicitando a los/as psicólogos/as que estén formados en ellos para aplicarlos. El autor de esta investigación advierte que la masificación del uso de este tipo de test y las malas prácticas en esta área se relacionan con deficiencias a nivel de la formación profesional e inexistencia de mecanismos de control de la calidad de esta formación, ausencia de regulación del ejercicio profesional por parte de organismos como el Colegio de Psicólogas y Psicólogos de Chile y pocos espacios de difusión de investigación especializada en el área (Didier, 2014).

Por otra parte, en el estudio de Rodríguez et al. (2017) se establece que en Chile la desregulación de empresas consultoras y la baja preparación profesional han propiciado la aparición de malas prácticas, dentro de las cuales se encuentra la utilización de instrumentos de evaluación de baja calidad técnica, es decir, sin pertinencia, fiabilidad o validez científica. Los/as autores/as señalan que las malas prácticas en este contexto tienen costos relevantes tanto para la organización, como para los/as postulantes y también para los/as psicólogos/as que deben realizar los procesos de selección. Sobre esto último, destacan que para los/as psicólogos/as, ser testigos o parte de prácticas inadecuadas aumenta la probabilidad de experimentar diversas formas de estrés, especialmente el distrés moral, aspecto que ha sido escasamente estudiado en el país.

Más recientemente, el único estudio en Chile que ha abordado la percepción de los/as psicólogos/as en el uso de test, realizado por Vinet et al. (2023) reporta el empleo habitual de test que no cuentan con fundamentación empírica y/o pertinencia para la población nacional. En este estudio se identificó que, dentro de los 20 test más frecuentemente utilizados, 13 corresponden a instrumentos de tipo proyectivo. Además, y con el objetivo de explorar la calidad psicométrica de los test más usados, a través de una revisión sistemática encontraron que de los 20, sólo cinco contaban con alguna investigación sobre su calidad psicométrica en población nacional (evidencias de confiabilidad, validez y datos normativos), todos ellos de tipo psicométrico, ninguno proyectivo.

Considerando este escenario nacional y el gran impacto que tienen las prácticas inapropiadas en el uso de test sobre la vida de las personas, se considera necesario instalar en Chile, de manera urgente, mecanismos que resguarden la formación y su adecuado uso, además de impulsar la investigación en psicometría. Estos mecanismos pueden articularse dentro de una política nacional que fomente la regulación del uso de test psicológicos y frene las malas prácticas, adhiriendo a los principios de la ética profesional y

garantizando el profundo respeto que merecen todas las personas evaluadas. La creación de mecanismos de regulación claros y efectivos podría garantizar mayor calidad, equidad, objetividad y seguridad en los procesos de evaluación psicológica en Chile.

## 2. Objetivo

El documento que se presenta a continuación presenta una revisión de las principales directrices que se han desplegado internacionalmente sobre el adecuado uso de instrumentos de evaluación psicológica, que se pueden vincular a los tres grandes pilares que se mencionaron anteriormente, a fin de identificar aquellos aspectos que son mínimamente necesarios para comenzar a desarrollar una política de buenas prácticas en el uso de test en Chile.

## 3. Método

Se realizó una revisión bibliográfica narrativa descriptiva. Esta metodología consiste en la presentación de una síntesis actualizada de diversas fuentes de información sobre un tema de investigación específico, acompañada de los comentarios, recomendaciones o conclusiones de los/as autores/as de la revisión, que se basan en su experiencia y conocimiento (Ato et al., 2013; Reyes, 2020; Pardal-Refoyo, 2023; Vera, 2009). Las fuentes de información seleccionadas para esta revisión incluyen artículos científicos y guías o manuales sobre orientaciones técnicas específicas en torno al adecuado uso de test, elaboradas por organismos especializados en medición y evaluación en psicología que cuentan con experiencia, reconocimiento y validación internacional en el área.

Como primer paso, se definió una estrategia de búsqueda en las bases de datos electrónicas de mayor relevancia en las Ciencias Sociales, tales como Web of Science, Scopus y Google Scholar. Adicionalmente se consideró la base de datos SciELO y páginas web de la APA e ITC, libros, manuales y guías técnicas de organismos internacionales especializados en el tema. Las palabras clave que guiaron la búsqueda de artículos científicos fueron “test”, “uso de test”, “directrices”, “buenas prácticas” y “práctica profesional”, en idioma español. Como segundo paso, se definieron los criterios de selección de las fuentes de información considerando su pertinencia para el abordaje del objetivo de la revisión (Guirao-Goris et al., 2008). De acuerdo con esto, se seleccionaron fuentes que: (a) entreguen directrices sobre el adecuado uso de test, (b) describan esfuerzos colectivos regulatorios sobre el uso de test ejecutados y evaluados en diversos países y (c) cuenten con reconocimiento o evidencia de consensos de la comunidad científica en torno al uso de test. De acuerdo al cumplimiento de estos criterios, las fuentes de información que se seleccionaron para organizar la propuesta de orientaciones se presentan en la Tabla 1.

Finalmente, y de acuerdo con esta metodología, luego de la búsqueda, revisión y selección de las fuentes, se clasificó la información en formato de fichajes, los cuales consisten en sistemas de organización que permiten ordenar y jerarquizar la información recopilada, establecer su relevancia y generar un análisis secuencial para su posterior inclusión en la escritura del texto (Loayza, 2021). De esta forma, esta revisión bibliográfica narrativa presenta una síntesis actualizada de las principales directrices internacionales especializadas en el tema, incluyendo conclusiones y recomendaciones por parte de la autora de este trabajo, con el fin de ofrecer lineamientos para el diseño de una política de buenas prácticas en el uso de test psicológicos.

**Tabla 1**
*Fuentes de información utilizadas para organizar la propuesta de orientaciones*

Tipo	Fuente (autor, año, título)	Síntesis de su contribución
Guías técnicas	ITC. (2013). International Test Commission Guidelines on Test Use.	Conjunto de orientaciones en torno al buen uso de test, elaborado por la comisión internacional que trabaja en medición psicológica, educativa y ocupacional, promoviendo prácticas, directrices y políticas justas, válidas, transparentes y eficientes.
	APA. (2020). APA Guidelines for Psychological Assessment and Evaluation.	Conjunto de directrices sobre evaluación psicológica de una organización científica y profesional que representa a la psicología en Estados Unidos, y que promueve el avance, la comunicación y la aplicación de la ciencia y el conocimiento psicológico.
Manual	AERA et al. (2018). Estándares para pruebas educativas y psicológicas (M. Lieve, Trad.). American Educational Research Association.	Manual publicado en colaboración por las tres organizaciones desde 1966, que representa el “gold standard” internacional en la orientación sobre pruebas.
Libros	Fernández-Ballesteros et al. (2011). Buenas prácticas y competencias en evaluación psicológica. El Sistema Interactivo Multimedia de Aprendizaje del Proceso de Evaluación (SIMAPE).	Libro que describe un sistema de multimedia para el entrenamiento en competencias técnicas necesarias para realizar un proceso de evaluación psicológica ético y basado en el método científico.
	Vinet & Rodríguez-Cancino. (2024). Evaluación Psicológica en Chile: Historia, Actualidad y Desafíos.	Capítulo de libro que discute la historia de la evaluación psicológica en Chile, las dificultades observadas y las necesidades para mejorar su ejercicio.
Artículos científicos	Elosua. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en psicología.	Artículo que aborda proyectos relacionados con el uso de test y la importancia de las comisiones y las directrices para mejorar su uso.
	Evers et al. (2012). Testing Practices in the 21st Century: Developments and European Psychologists’ Opinions.	Artículo que resume acciones y proyectos de la Federación Europea de Asociaciones de Psicólogos (EPPA) para mejorar las prácticas de evaluación en los países europeos.
	Fernández-Ballesteros et al. (2003). Guías para el proceso de evaluación (GAP): Una propuesta a discusión.	Artículo que describe el desarrollo de las Guías del Proceso de Evaluación (GAP), patrocinado por la Asociación Europea de Evaluación Psicológica (EAPA) como una propuesta sobre la necesidad de mejorar la práctica de la evaluación psicológica y la formación en el área.
	Hernández et al. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test.	Artículo que revisa las principales actividades y proyectos llevados a cabo por la Comisión de Test del Consejo General de Psicología de España para mejorar la calidad y el uso de los instrumentos de evaluación.
	Muñiz et al. (2011). Evaluación de test editados en España.	Artículo que reporta las primeras evaluaciones de test realizadas a través de un modelo de evaluación desarrollado por la Comisión Europea de Test en el contexto español.
	Muñiz et al. (2020). El uso de los Test en España y Latinoamérica: Perspectivas actuales y retos futuros.	Artículo que ofrece una visión general y actualizada sobre las prácticas en el uso de test tanto en España como en Latinoamérica, y sus desafíos futuros
	Prieto & Muñiz. (2000). Un modelo para evaluar la calidad de los tests utilizados en España.	Artículo que describe un modelo de evaluación y difusión de los aspectos técnicos de los test como un medio para mejorar su uso, a cargo del Colegio Oficial de Psicólogos de España.
	Vinet et al. (2023). El Empleo de Test por Psicólogos/as Chilenos/as: Un Inquietante Panorama.	Artículo que presenta los resultados del único estudio (a la fecha) en Chile que ha abordado el uso de test, la evidencia de su calidad psicométrica y propuestas para mejorar las buenas prácticas.
	Wechsler et al. (2019). O desenvolvimento da avaliação psicológica no Brasil: Avanços históricos e desafios.	Artículo que aborda la historia de la evaluación psicológica en Brasil, identificando como hitos relevantes la fundación del Instituto Brasileño de Evaluación Psicológica y el Sistema de Evaluación de las Pruebas Psicológicas (SATEPSI) para la mejora en el uso de test.

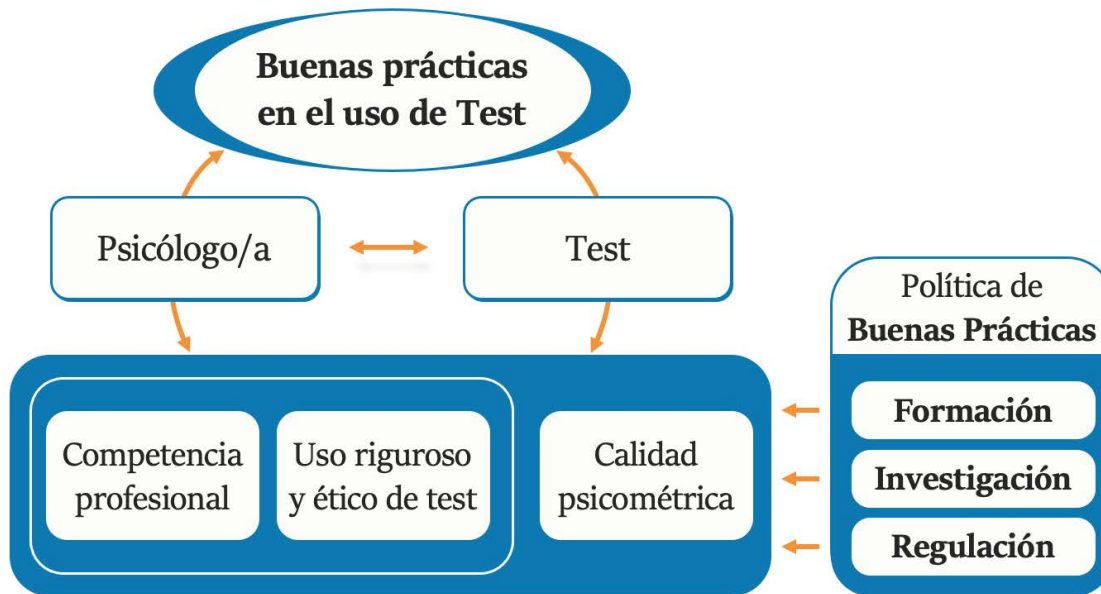
#### 4. Resultados

El desarrollo del tema que se presenta en esta revisión se estructura de acuerdo con los tres pilares esenciales antes mencionados: la competencia profesional y el uso ético y riguroso de los test, que ponen el foco en quien administra el instrumento; y la calidad psicométrica de los test, que pone el énfasis en el instrumento mismo. Para cada uno de estos ejes, se incluyen consideraciones para una política de buenas prácticas en el uso de test en el contexto chileno, las que se desarrollan en la sección de recomendaciones para la acción. La Figura 1 ilustra y sintetiza el abordaje del tema.



**Figura 1**

*Esquema de buenas prácticas en el uso de test psicológicos*



Fuente. Elaboración propia

#### **4.1 Competencia profesional: Formación sólida y pertinente en el uso de test**

El rápido y continuo desarrollo de instrumentos, procedimientos, avances normativos, tecnología y la evolución de las prácticas basadas en la evidencia, obliga a los/as psicólogos/as a buscar capacitación permanente en sus procedimientos evaluativos (APA, 2020). Los test se van actualizando permanentemente y la evidencia empírica va generando información que se debe incorporar en su uso, por lo que es fundamental que el/la psicólogo/a se capacite, entrene y supervise su quehacer, con pares o personas que posean mayor experiencia en el uso de algún test específico.

La competencia en el uso de test se refiere al desarrollo de un sólido conocimiento de sus principios teóricos y empíricos. Esto significa que el/la psicólogo/a debe ser experto/a en los procedimientos adecuados de administración y corrección de los test con que trabaja, así como también en los fundamentos teóricos, empíricos y psicométricos que avalan la interpretación de sus puntuaciones.

Según Forns y Amador (2017) la adecuada administración de test e instrumentos de evaluación requiere que el/la psicólogo/a sea competente en:

- (1) habilidades cognitivas: conocimiento de los instrumentos de evaluación, su fundamentación teórica, constructos que evalúa y áreas de aplicación;
- (2) habilidades procedimentales: conocimientos sobre las normas y reglas de administración, puntuación, interpretación de resultados y entrenamiento específicos en las pruebas que lo requieren; y
- (3) habilidades clínicas: habilidades de observación del lenguaje verbal y no verbal de la persona evaluada durante la administración de los instrumentos y habilidad para relacionar su conducta con los resultados, con los fundamentos teóricos de los test y la posible presencia de alteraciones o disfunciones.

Por otra parte, las directrices de la APA (2020) indican que los/as psicólogos/as deben esforzarse por desarrollar y mantener habilidades para seleccionar, utilizar e interpretar test, integrar los hallazgos con otras fuentes de información que sean relevantes para el motivo de consulta y comunicar adecuadamente los resultados.

Los/as psicólogos/as deben asegurarse de que son competentes en el uso de test y ofrecer servicios sólo en aquellos en que cuenta con la preparación necesaria. Además, los/as psicólogos/as deben conocer los límites de su propia competencia y no actuar fuera de ellos. Esto implica la obligación de mantenerse actualizado/a en relación a los cambios y avances en la construcción y uso de test, además de conocer exhaustivamente los marcos legales que rigen su uso (Fernández-Ballesteros et al., 2011; ITC, 2013).

Algunos test demandan esfuerzos importantes a los/as evaluadores/as, tales como el rápido registro de las respuestas, manipulación rigurosa y estandarizada de materiales o equipos, o ejecución de ítems y entrega de instrucciones de alta complejidad. Ante estos, las directrices de la AERA et al. (2018) plantean que, si los/as examinadores/as no cuentan y/o no pueden desarrollar las habilidades necesarias para dar un correcto cumplimiento a estas exigencias, deben ser conscientes de sus limitaciones personales y no aplicar ese tipo de pruebas.

#### **4.2 Calidad de los test: valoración de sus propiedades psicométricas**

Los/as profesionales que utilizan pruebas educativas o psicológicas y que deben tomar decisiones relevantes con la información que obtienen de ellas, deben asegurarse de que sus evaluaciones cuentan con un respaldo científico que avale sus conclusiones y recomendaciones (Sireci & Benítez, 2023). En este ámbito contar con la información disponible sobre las propiedades psicométricas de los test (evidencias de validez, confiabilidad e imparcialidad) y verificar su calidad a través de ellas, puede garantizar que su uso sea apropiado y justo (AERA et al., 2018; ITC, 2013; Muñiz & Fonseca-Pedrero, 2019; Vinet et al., 2023).

##### **4.2.1 Evidencias de validez**

El concepto psicométrico de validez se refiere al grado en que la evidencia y la teoría respaldan la interpretación prevista de las puntuaciones de un test para el uso propuesto (AERA et al., 2018). Esta definición asume que la validez no es una propiedad inherente al test, y que ésta se refiere al uso propuesto y a la interpretación de sus puntuaciones, por lo que éstas podrían ser válidas para un propósito (contar con soporte empírico y teórico), pero inválidas para otro (Elosua, 2012; Leong et al., 2020; Sireci & Benítez, 2023).

El proceso de obtención de evidencias de validez de un test consiste en la recopilación y análisis de pruebas que permitan defender su uso para un fin determinado e interpretar sus puntuaciones (Elosua, 2017; Sireci, 2020). Leong et al. (2020) señalan que la validación es un proceso de construcción y evaluación de evidencias a favor o en contra de la interpretación de las puntuaciones de un test para un determinado propósito. Este proceso, que comienza incluso antes de la construcción del test, y que orienta su desarrollo y vigencia, incluye un conjunto de estudios que permitan otorgar una interpretación teórica coherente a sus puntuaciones (Elosua, 2003; Muñiz & Fonseca-Pedrero, 2019). La revisión de las evidencias recopiladas en los procesos de validación son especialmente relevantes para un/a psicólogo/a cuando debe seleccionar un test para un fin específico (Vinet et al., 2023).

Cabe destacar que de acuerdo con las directrices de la AERA et al. (2018) la validación es una responsabilidad compartida entre el/la desarrollador/a de un test y el/la profesional que lo utiliza. Por una parte, quien desarrolla el instrumento debe recopilar y proveer evidencia relevante que fundamente cualquier interpretación de sus puntajes en el marco del propósito previsto y por otra,

el/la psicólogo/a que decide usarlo debe ser capaz de evaluar la pertinencia de esta evidencia en el contexto particular en que utilizará el test.

Diversos autores (Brenlla et al., 2023; Elosua, 2003; Iliescu et al., 2024; Leong et al., 2020; Sireci & Benítez, 2023) coinciden con la clasificación propuesta por la AERA et al. (2018) que señala que las fuentes de evidencias de validez se basan en el contenido del test, su estructura interna, su relación con otras variables, el proceso de respuesta o las consecuencias de su uso.

La Tabla 2 describe de manera sencilla algunos ejemplos del alcance y naturaleza de cada una de las fuentes de evidencia de validez.

**Tabla 2**

*Síntesis de las fuentes de evidencias de validez consideradas en la definición actual de validez*

Fuentes de evidencia de validez basadas en:	Permite verificar si:	Ejemplos
Contenido	Los ítems de un test <b>representan adecuadamente la diversidad de aspectos</b> que posee aquello (constructo) que se pretende medir.	Un test de conocimientos sobre las cuatro operaciones aritméticas cuyos ítems indagan suma, resta, multiplicación y división (alta validez de contenido) versus otro test cuyos ítems indagan solo suma y resta (baja validez de contenido).
Estructura interna	Los ítems y subescalas de un test confirman el <b>patrón de asociaciones internas teóricamente esperado</b> .	Los ítems de un test de empatía se organizan empíricamente en dos dimensiones teóricamente identificables como empatía cognitiva y empatía afectiva (alta validez estructural) versus se organizan en una dimensión única (baja validez estructural).
Relación con otras variables	Los puntajes de un test se <b>asocian del modo teóricamente esperado</b> con (a) otras medidas del mismo constructo (validez concurrente), (b) medidas de otros constructos (validez convergente) o (c) no se asocian con medidas de otros constructos (validez discriminante).	Los puntajes de un test de depresión muestran una correlación positiva con otro test de depresión (alta validez concurrente), una correlación positiva con un test de ansiedad o negativa con un test de optimismo (validez convergente), o muestran ausencia de correlación con un test de deseabilidad social (validez discriminante).
Proceso de respuesta	Las respuestas a los ítems del test <b>activan realmente los procesos psicológicos</b> que teóricamente deberían activarse.	Un test que mide apoyo familiar supone que todos los respondientes comparten el mismo concepto de familia (personas con relaciones parentesco). Si las personas que viven solas incluyen dentro de su familia a conocidos o vecinos, entonces ese test tiene una baja validez de proceso de respuesta.
Consecuencias de su uso	La interpretación o uso de los resultados de un test <b>tiene consecuencias intencionadas o no intencionadas</b> .	Una evaluación de la excelencia pedagógica genera en los profesores favorecidos una mayor autoestima profesional (consecuencia intencionada: alta validez), pero escaso reconocimiento de sus pares, directivos y sostenedores (consecuencia no intencionada: baja validez).

#### 4.2.2 Confiabilidad/precisión de la medida

Todo profesional que utilice los test como una de las fuentes de información que orientarán la toma de decisiones, debe asegurarse de que sus puntuaciones sean razonablemente confiables, es decir, consistentes y precisas (Urbina, 2014). La confiabilidad es una condición fundamental de un instrumento de medición y un indicador de su calidad. Si un test posee una baja confiabilidad sus puntajes serán inestables o inconsistentes reflejando fundamentalmente el error de medición (Mena, 2019; Iliescu et al., 2024).

La confiabilidad no es una cuestión de “todo o nada”, si no de grado, siendo más recomendable visualizarla como un continuo que puede ir desde la consistencia mínima de una medición hasta la casi perfecta replicación de los resultados (Gregory, 2012). Los/as profesionales que utilizan test deberían examinar esta propiedad y seleccionar sólo aquellos que cuenten al menos con un nivel aceptable de confiabilidad.

El examen del grado de confiabilidad de un test es uno de los elementos que permite garantizar que sus puntajes son útiles, es decir, lo suficientemente consistentes y libres de errores de medición (Urbina, 2014), otorgando a el/la profesional la seguridad de que sus decisiones de diagnóstico o intervención cuentan con un fundamento psicométrico robusto. Es importante considerar que la confiabilidad/precisión puede variar de una población a otra, dado que las fuentes de error como los efectos de el/la evaluador/a, la familiaridad con los ítems o los materiales de un test, pueden tener mayor impacto en una población que en otra (AERA et al., 2018). Por lo tanto, no se debe asumir que, si la exploración psicométrica de un test arroja buenos índices de confiabilidad para una muestra de personas en particular, estos mantendrán su calidad si este se aplica a otras que difieren en sus características personales o culturales. Eso debe ser examinado y comprobado.

La confiabilidad/precisión de una medición depende de una serie de factores que tienen que ver con quien realiza la medición (examinador/a), con lo que utiliza para realizarla (instrumento de medición-test) y con lo que se está midiendo u observando (Mantenera et al., 2018). Gregory (2012) señala que las puntuaciones en un test siempre serán el resultado de factores que contribuyen a la consistencia (atributos estables que se pretenden medir) y de factores que contribuyen a la inconsistencia (características de la persona evaluada, del test o la situación de evaluación), que no se relacionan con el atributo a medir, pero que afectan las puntuaciones. Según este autor esto se traduce en que la puntuación que obtiene una persona evaluada en un test (puntaje observado) siempre incorpora la medida del atributo, más los elementos de error de medición, haciendo que la puntuación verdadera (sin errores), nunca pueda ser conocida. Por lo tanto, las puntuaciones de un test solo representan una estimación de la puntuación verdadera del atributo que se esté midiendo, la cual tenderá a ser mejor, en la medida que la medición incremente su fiabilidad.

En el contexto del uso de test, el error de medición se puede definir como cualquier fluctuación en las puntuaciones de un test que resulta de factores presentes en el proceso de evaluación, pero que son irrelevantes para el atributo que se mide. Al revisar las propiedades psicométricas de los test y su grado de confiabilidad, es importante recordar que estas fuentes de error de medición reducen la utilidad de los puntajes de un test y la posibilidad de confiar en la precisión de los resultados, por tanto, la confiabilidad y el error de medición son dos aspectos interdependientes y estrechamente vinculados a qué tan consistente es un instrumento de medida (AERA et al., 2018; Gregory, 2012; Urbina, 2014).

#### **4.2.3 Imparcialidad**

Esta propiedad psicométrica puede ser entendida como la sensibilidad para considerar respetuosamente todas las características de las personas evaluadas (como etnia, género, edad, nivel socioeconómico, contexto lingüístico o cultural), durante todas las etapas de desarrollo, aplicación, calificación, interpretación y uso de los test psicológicos (Díaz-López et al., 2017). Un test en particular no necesariamente es capaz de medir los mismos constructos en personas que pertenecen a distintos subgrupos, puesto que el contenido o el formato del instrumento puede impactar diferencialmente los puntajes de los/as examinandos/as, favoreciendo o desfavoreciéndolos involuntariamente. Por lo tanto, si se decide utilizar un test en un subgrupo en particular, se debe realizar un análisis riguroso de la evidencia de validación del instrumento existente para dicho subgrupo (AERA et al., 2018; Brenlla et al., 2023).

Una prueba es imparcial si refleja los mismos constructos para todas las personas evaluadas, sin favorecer o perjudicar a alguien debido a sus características personales o pertenencia a un grupo específico de la población. Cuando esto no ocurre, aparece una amenaza central a la imparcialidad: el sesgo de medición (AERA et al., 2018). Frente al tema del manejo de los posibles sesgos en los

test, los lineamientos de la ITC (2013) establecen que si un test se utiliza con personas de diferentes grupos (según género, cultura, nivel educacional, etnia, etc.) es responsabilidad de el/la psicólogo/a garantizar que son imparciales y adecuados para ellos. Para esto el/la profesional debe asegurarse de que los constructos que se están midiendo son relevantes para cada uno de los grupos y debe ser capaz de recopilar y revisar la evidencia en torno al funcionamiento diferencial de los ítems y/o los procesos de validación que apoyen su uso en diferentes grupos.

Asimismo, estos lineamientos señalan que cuando se utilicen test en más de un idioma es necesario asegurarse que en la construcción de sus distintas versiones se hayan considerado rigurosamente todos los aspectos de contenido, culturales e idiomáticos, y que quien los aplique sea capaz de comunicarse perfectamente en el idioma en que se debe administrar el test.

#### **4.3 Uso ético y riguroso de los test**

La utilización adecuada de los test implica que el/la psicólogo/a sea capaz de seleccionarlos, administrarlos, corregirlos e interpretarlos de manera correcta y responsable. Esto también incluye que el/la profesional garantice su cuidado y proteja su integridad respetando su copyright (prohibición sobre la copia), evitando la filtración o publicación de sus ítems en medios de comunicación masivos o redes sociales, y resguarde que los instrumentos sean utilizados solo por personas que cuenten con la debida capacitación y habilitación necesaria para ello (Hall et al., 2005; ITC, 2013). Estos aspectos que se refieren a la seguridad de los test, son fundamentales para garantizar la validez de las interpretaciones de sus puntuaciones. Por ejemplo, se ha observado un aumento en la disponibilidad de tecnologías para fotografiar o compartir registros de los ítems de los test, o los intentos de conseguirse las pruebas antes de responderlas. Esto arriesga la potencial utilidad de la medición, ya que no estará reflejando correctamente el constructo que pretende medir, lo que puede tener serias consecuencias para las personas evaluadas y para la toma de decisiones (Muñiz et al., 2015).

##### **4.3.1 Selección de test**

La selección de los test debe realizarse de acuerdo al motivo de consulta, los objetivos de la evaluación y las hipótesis elaboradas durante el proceso evaluativo (Forns & Amador, 2017). En esta tarea es necesario utilizar preferentemente herramientas de evaluación cuya calidad psicométrica sea sólida, es decir, que cuenten con evidencia de validez para su uso, suficiente confiabilidad de la puntuación y capacidad para generar medidas que sean justas y apropiadas para el propósito de la evaluación, la población, el entorno y el contexto en cuestión (APA, 2020).

Según las directrices de la ITC (2013) los/as psicólogos/as deben evaluar la potencial utilidad de los test para cada situación evaluativa, justificando razonablemente su uso y solo luego de ejecutar un análisis riguroso de las necesidades y características de la persona que evaluará, además de considerar siempre fuentes adicionales de información. El/la profesional deberá sopesar las ventajas e inconvenientes de escoger algún test en particular frente a otras fuentes de información.

Por otra parte, la ITC (2013) indica que el/la psicólogo/a debe escoger test que sean técnicamente adecuados y pertinentes a cada situación evaluativa. Para esto, debe examinar toda la información disponible y comprobar que la documentación técnica sobre el test incluye información sobre la representatividad de su contenido, características de los grupos normativos utilizados en la baremación, confiabilidad/precisión de la medida, evidencias de validez y ausencia de sesgos que respalden su utilización. El/la profesional no debería utilizar instrumentos que tengan una documentación técnica inadecuada o poco clara, y no deberían aceptar un test

basándose únicamente en su validez aparente, es decir, la ilusión de que algo es útil o relevante por el solo hecho de que es ampliamente utilizado, o recomendado por otros/as usuarios/as, o por consejos de quienes tienen intereses comerciales.

En la selección de test es esencial considerar las evidencias de validez y confiabilidad/precisión descritas anteriormente, así como también las características demográficas de los grupos para los que el test fue diseñado originalmente. Seleccionar test que cuenten con normas demográficas y clínicamente pertinentes para la persona evaluada y para el objetivo de la evaluación, permite respaldar las inferencias que se realicen tras su administración. No resulta apropiado aplicar, a otros grupos, una prueba construida para un grupo específico sin un previo examen de su pertinencia (AERA et al., 2018).

#### 4.3.2 Administración de test

Una correcta administración de test psicológicos demanda del desarrollo de habilidades cognitivas, procedimentales y clínicas por parte de el/la evaluador/a (Forns & Amador, 2017), además de la necesidad de asegurarse que existan adecuadas condiciones internas y externas a la persona evaluada.

##### 4.3.2.1 Rapport

De acuerdo con las directrices de la ITC (2013) aplicar los test de manera adecuada primeramente requiere que el/la evaluador/a establezca una buena relación con las personas evaluadas, dirigiéndose a ellas de forma positiva e intentando reducir su ansiedad ante la situación de evaluación. Además, es necesario que considere la eliminación de potenciales fuentes de distracción, tales como alarmas de relojes o teléfonos móviles, y se asegure de que dispone de todos los materiales necesarios para la aplicación del test, antes de comenzar con su administración.

Generar una buena relación con la persona evaluada implica saber establecer *rapport*, concepto que Forns y Amador (2017) definen como la habilidad para mantener una escucha activa, ser capaz de adaptarse a las características específicas de la persona evaluada, respetar su *timing* o ritmo, entre otras. El establecimiento del *rapport* puede considerarse como una habilidad crítica para los/as examinadores/as puesto que se ha evidenciado que la calidad de la relación puede aumentar o disminuir el desempeño de las personas en pruebas estandarizadas, lo que puede afectar significativamente las decisiones profesionales y, por ende, tener efectos relevantes en la vida de las personas (Hall et al., 2005).

##### 4.3.2.2 Observación del desempeño

Otro elemento de gran relevancia en la administración de test es la capacidad de observación y registro de el/la evaluador/a. Observar y describir la conducta verbal y no verbal de la persona evaluada, su actitud hacia la situación de evaluación, sus estrategias de resolución de problemas, disposición afectiva, nivel de atención, estado de ánimo, o sus habilidades sensoriales y motrices, son un complemento indispensable para la interpretación de las puntuaciones que emergen de un test (Cortés & Benavente, 2007; Echavarría-Ramírez & Tirapu-Ustárrroz, 2021; Rodríguez-Cancino et al., 2024; Sattler, 2010).

Los/as expertos/as en procesos de evaluación psicológica han subrayado la importancia de mirar más allá de las puntuaciones de los test. Una de las principales ventajas de los test psicológicos que se administran de manera individual es la posibilidad de observar detenidamente el desempeño de la persona evaluada permitiendo a el/la examinador/a determinar si las puntuaciones obtenidas reflejan de manera precisa la capacidad de una persona frente a una tarea estandarizada. El desarrollo de esta habilidad requiere de un sólido entrenamiento (Hall et al., 2005; Rodríguez-Cancino et al., 2024).

#### 4.3.2.3 *Condiciones ambientales*

En cuanto a las condiciones externas para la adecuada administración de pruebas, las directrices de la AERA et al. (2018) señalan que es responsabilidad de quienes desarrollan los test proporcionar toda la información sobre las condiciones del entorno que se requieren para aplicar un instrumento, mientras que es responsabilidad de el/la evaluador/a garantizar su cumplimiento. Además, destacan explícitamente que no se puede lograr una adecuada medida de un constructo si el espacio de trabajo es inadecuado, ruidoso, existe una inadecuada temperatura y luminosidad o se usan materiales de baja calidad o ilegibles.

#### 4.3.2.4 *Procedimientos estandarizados*

Otro aspecto de gran importancia en la adecuada administración de test es el riguroso cumplimiento de los procedimientos estandarizados. Las pruebas estandarizadas utilizan los mismos materiales, reglas de administración y de corrección a todas las personas evaluadas. La importancia de adherirse rigurosamente a estas condiciones y no alterarlas, radica en que esto es un mecanismo que promueve la imparcialidad, facilita las comparaciones de puntajes entre individuos y garantiza el uso de los datos normativos. En los test que han sido diseñados para evaluar los conocimientos, habilidades, capacidades u otras características personales, el uso de procedimientos estandarizados permite garantizar que todas las personas evaluadas cuenten con las mismas oportunidades de demostrar sus competencias, y nadie tenga una ventaja indebida (AERA et al., 2018).

El incumplimiento o la desviación de los procedimientos estandarizados, en teoría, disminuyen la fiabilidad y validez de los resultados de la prueba, ya que la precisión de la medida que se obtiene tras la aplicación de un test no sólo reside en el instrumento, sino que también depende de las circunstancias de su uso (Hall et al., 2005). Si existen ocasiones en que se hace necesario ajustar o modificar los procedimientos estandarizados de administración de un test, el/la profesional debe ser consciente de que cualquier modificación producirá cambios en la estimación del constructo que originalmente medía el instrumento. Por esta razón, cualquier modificación que cambie el constructo, invalidará inmediatamente la posibilidad de usar las normas para las interpretaciones de esos puntajes (AERA et al., 2018). En estos casos, antes de utilizar un test modificado de su versión original se debe reunir toda evidencia que sea necesaria para estimar la validez de las interpretaciones para los usos previstos de los puntajes.

#### **4.3.3 Interpretación y comunicación de resultados de los test**

Luego de una adecuada selección, administración y corrección de test, el/la profesional se adentra en la compleja tarea de interpretar sus resultados, integrando todas las fuentes de información relevante para luego comunicar los resultados. Este proceso tiene una alta relevancia para la persona evaluada ya que sobre dicha información se tomarán decisiones de derivación, intervención y/o planificación de tratamientos. En esta tarea es necesario que el/la psicólogo/a integre, en una mirada global y comprensiva, toda la información recopilada en el proceso de evaluación junto a sus conocimientos de psicología, psicopatología y psicometría, y despliegue apropiadas destrezas lingüísticas y comunicativas (Cortés & Benavente, 2007; Fernández-Ballesteros et al., 2011; Forns & Amador, 2017; Rodríguez-Cancino et al., 2024).

De acuerdo con las directrices de la AERA et al. (2018) los manuales de los test indican las variables que deberían tomarse en cuenta al interpretar las puntuaciones, destacando los antecedentes clínicamente relevantes, uso de medicamentos, antecedentes escolares, laborales o vocacionales de la persona evaluada, entre otros. Además, destacan que los antecedentes relativos a la edad, cultura, discapacidad, género y características lingüísticas o raciales/étnicas siempre deben ser considerados de manera relevante.



Por su parte, los lineamientos de la ITC (2013) indican que para que el/la psicólogo/a pueda interpretar adecuadamente los resultados de un test, debe contar con una profunda comprensión de sus fundamentos teórico/conceptuales, de los tipos de puntuaciones utilizadas y de las normas, así como también de las limitaciones de los instrumentos de medida que haya administrado. En la tarea de interpretación, el/la profesional debe intentar minimizar cualquier posibilidad de sesgo y considerar toda la información disponible sobre la persona evaluada (*p. ej.*, edad, género, escolaridad o cultura), evitando perpetuar estereotipos o realizar generalizaciones de los resultados a rasgos o características de la persona que no han sido medidos por el test. También se destaca que en la interpretación de las puntuaciones se debe considerar las evidencias de validez, la fiabilidad y el error de medición, y todos aquellos elementos que puedan alterar artificialmente los resultados.

Cuando se realicen reportes (orales o escritos) de las interpretaciones de las puntuaciones de los test, se debe colocar énfasis en el uso apropiado de esta información y en minimizar las potenciales consecuencias negativas que se deriven de ésta. Se sugiere que el reporte de la precisión de las puntuaciones se acompañe de los márgenes de error o rangos probables y se incluyan referencias sobre todos los factores que podrían afectar a los resultados e interpretaciones. Es responsabilidad de el/la psicólogo/a evitar las interpretaciones erróneas y velar para que no se haga un uso indebido de los resultados o las puntuaciones (AERA et al., 2018).

En cuanto a la comunicación de los resultados, la ITC (2013) indica que el/la psicólogo/a debe ser capaz de realizarlo de forma clara y precisa, ya sea de manera oral o escrita, ajustada a las características, intereses y roles de el/la receptor/a de estos. También señala que el/la profesional debe ser capaz de resguardar la confidencialidad de los datos e identificar a qué personas o instituciones es relevante y pertinente compartir los resultados de un proceso de evaluación, siempre con el explícito consentimiento de la persona evaluada.

## 5. Recomendaciones para la acción

### 5.1 Consideraciones para la formación en uso de test en Chile

De acuerdo con estos antecedentes y el escenario actual ilustrado por Vinet y Rodríguez-Cancino (2024), una política que promueva las buenas prácticas en el uso de test en Chile debe considerar el fortalecimiento en la formación de pregrado y la necesidad de crear programas de postítulo y postgrado que permitan a los/las profesionales el adecuado desarrollo de estas competencias profesionales.

El fortalecimiento de la formación inicial requiere que las universidades que imparten las carreras de psicología acuerden estándares y contenidos mínimos, basados en las directrices internacionales y en las necesidades de formación nacionales. Al respecto, es mínimamente necesario y obligatorio incluir una adecuada formación en Psicometría que les permita a los/as profesionales escoger adecuadamente los test que son más pertinentes frente a un motivo de consulta en particular; formación en habilidades clínicas para administrar correctamente los instrumentos; y consolidación de un marco conceptual robusto (por ejemplo en procesos afectivos y cognitivos, desarrollo humano, trayectorias normativas y no normativas, psicopatología, etc.), que permita a los/as psicólogos/as desplegar las habilidades de interpretación e integración de los resultados de distintos instrumentos de medida.

Por otra parte, en línea con los planteamientos de Juliá (2023) debería ser obligatorio que la formación profesional inicial promueva enfáticamente la necesidad y el valor de la reflexión ética constante sobre las prácticas en el ejercicio profesional, propiciando que se complemente la práctica basada en la evidencia, con la reflexión permanente sobre las decisiones que toman y las acciones que realizan los/as psicólogos/as.

A nivel de postítulo o postgrado, y dada la obligatoriedad de actualización permanente, sería necesario contar con instancias de especialización en áreas de evaluación o test específicos. La profundización en Psicometría y la supervisión de procesos de evaluación psicológica, en que se utilicen test, son ejemplos de ello, siempre que cuenten con mecanismos que garanticen la calidad de la formación que imparten. En este sentido, todas aquellas instancias asociadas a las universidades que cuentan con mecanismos de aseguramiento de la calidad, ya sea en la formación de postgrado o de educación continua, serían el espacio más confiable e idóneo para el desarrollo de esta propuesta.

La formación en el adecuado uso de test, tanto a nivel de pregrado como de postgrado, también debería incluir el fortalecimiento de las competencias de el/la psicólogo/a para valorar la calidad psicométrica de los instrumentos de medida que escoge utilizar dentro de un proceso de evaluación psicológica. Esto implica que, al menos, desarrolle la habilidad de comprender qué significan e interpretar adecuadamente las evidencias de validez, confiabilidad e imparcialidad. Sólo así, en el momento de seleccionar un test podrá determinar, fundamentadamente, si le es útil como instrumento de medida para un motivo de evaluación en particular.

### **5.2 Consideraciones para la investigación y estudio de la calidad de los test en Chile**

Las propiedades psicométricas de los test permiten certificar su calidad como herramientas de medición idóneas, por lo que es necesario estar constantemente evaluándolas y así garantizar su uso en una población en particular. Tal como se mencionó anteriormente, en Chile la investigación y desarrollo de instrumentos se ha visto paulatinamente fortalecida, sin embargo, para que esto no se estanque, es necesario contar con políticas nacionales que promuevan su consolidación (Vinet et al., 2023).

El estudio de Vinet et al. (2023) evidenció que los/as psicólogos/as chilenos/as utilizan ampliamente test proyectivos que no cuentan con evidencia sobre sus propiedades psicométricas y/o pertinencia cultural. Por otra parte, encontraron que de los 20 test más utilizados a nivel nacional, sólo cinco de ellos contaban con este tipo de evidencia. Por lo tanto, la investigación en psicometría en Chile debiese estar orientada a la evaluación constante de la calidad de la totalidad de los test que se usan masivamente, en términos de su precisión, pertinencia e imparcialidad, garantizando su utilidad para los usos propuestos.

En esta línea, se considera necesario que las instituciones y organizaciones validadas para dicha tarea (como las universidades, centros o sociedades científicas), en constante articulación con las que se relacionan con el ejercicio profesional (*p. ej.*, el Colegio de Psicólogas y Psicólogos) velen por mantener o crear equipos de investigación especializados en generar y difundir evidencia sobre las propiedades psicométricas (evidencias de validez, confiabilidad e imparcialidad) de los test que se utilizan frecuentemente en la práctica profesional, a fin de respaldar o frenar su uso masivo.

Esta tarea, además de promover la investigación sobre las propiedades psicométricas de los test, debería incluir la instalación de mecanismos que puedan valorar su calidad, garantizando que su uso sea justo y pertinente para la población chilena, especialmente en aquellos instrumentos en los que se basa la toma de decisiones que impactan profundamente la vida de las personas, ya sea en el ámbito clínico (diagnóstico o tratamiento), educacional (ingreso a programas de apoyo específicos), laboral (selección de personas)

o judicial (consejería o peritajes). En este sentido, sería recomendable seguir el camino recorrido por países como España (Hernández et al., 2022; Prieto & Muñiz, 2000) o Brasil (Wechsler et al., 2019; Wechsler et al., 2022) quienes han instalado sistemas de valoración de los test que certifican su calidad técnico-científica de acuerdo con el nivel de cumplimiento de exigencias teóricas y psicométricas mínimas, que permitan aprobar o restringir su uso en el campo profesional.

Es importante señalar que en este ámbito también cumplen un rol fundamental quienes se encargan de la edición y/o comercialización de los test. En este sentido, una política de buenas prácticas en el uso de instrumentos en Chile debiera incluir estrategias que permitan exigir y fiscalizar si dichas entidades toman en consideración la valoración de la calidad psicométrica de los instrumentos y cuentan con mecanismos que permitan garantizar su pertinencia para la realidad local.

Por último, aquí también es indispensable que la formación de pregrado y postgrado esté alineada con los hallazgos de la investigación en psicometría. Dado que se ha observado un uso extendido de pruebas proyectivas sin evidencia sobre su pertinencia para la población chilena, es posible suponer que los/las psicólogos/as reciben formación en ellas. Por lo tanto, sería recomendable que los planes y programas de formación seleccionen de manera rigurosa los test que van a enseñar, privilegiando aquellos que cuentan con una valoración de su calidad psicométrica, lo suficientemente adecuada para garantizar su utilidad.

A pesar de que se haya demostrado que las propiedades psicométricas de un test son adecuadas, éste puede perder su eficacia como herramienta de medición si lo utiliza alguien que no cuente con las competencias necesarias para ello (Muñiz et al., 2011). También puede perder su eficacia si no se controlan las fuentes de error de medición (en la medida de lo posible), o si el/la psicólogo/a no lo utiliza de manera ética y rigurosa.

### ***5.3 Consideraciones para la regulación del uso ético y riguroso de los test en Chile***

La evidencia internacional ha mostrado que el uso ético y riguroso de los test requiere la existencia de mecanismos de regulación de la práctica profesional. Sin embargo, tal como señalan Vinet y Rodríguez-Cancino (2024) estos mecanismos actualmente no existen y son altamente necesarios en Chile. Al respecto cabe destacar, que si bien es cierto hoy en día el Colegio de Psicólogas y Psicólogos no cuenta con la posibilidad de ejercer control sobre la práctica profesional, a nivel nacional se está avanzando en que esto cambie a partir de un proyecto de ley que pretende devolver esta facultad a todos los Colegios Profesionales (Carrillo, 2024). En el escenario de que esta iniciativa cuente con un resultado favorable sería esperable la implementación de medidas tales como la colegiatura obligatoria, acreditación de formación en test específicos y mecanismos de sanción ante las faltas éticas.

Cabe destacar que el uso ético y riguroso de los test no puede ampararse sólo en mecanismos de regulación externos. También es imprescindible que el/la psicólogo/a autorregule sus prácticas, sobre la base de un marco ético personal y profesional profundamente desarrollado. Es crucial que este aspecto sea una prioridad en la formación de las competencias profesionales iniciales.

Por otra parte, también contribuiría al uso ético y riguroso de los test, el examen de las evidencias de validez basadas en las consecuencias de su aplicación. Este proceso permitiría determinar qué impacto o implicancias ha tenido su uso para las personas evaluadas, pudiendo alertar sobre malas prácticas, utilización inadecuada de los resultados, entre otras.

El complemento de mecanismos de regulación internos y externos permitirían avanzar de manera importante en garantizar que los/as psicólogos/as cumplan con estándares mínimos para una adecuada selección, administración y corrección de los test, e

interpretación y comunicación de sus resultados. La elaboración de una política sobre buenas prácticas en el uso de test permitiría articular eficientemente todos los desafíos nacionales actuales a nivel de la formación, investigación y regulación del ejercicio profesional, anteriormente expuestos.

## 6. Limitaciones y futuras direcciones

Dentro de las limitaciones de este trabajo se encuentra que dada las características de una revisión narrativa la selección de las fuentes de información suele ser más subjetiva, dependiente de las perspectivas o intereses de sus autores/as, y por tanto puede incluir sesgos. Sería recomendable que se realice una revisión de las prácticas sobre el uso de test en Chile y/o Latinoamérica, a través de una metodología de revisión sistemática o metaanálisis, para contrarrestar eventuales fuentes de sesgos. También se sugiere dar continuidad al abordaje empírico de esta temática en la línea de lo realizado por el equipo de Vinet et al. (2023), a fin de identificar nuevas necesidades y retos. Los acelerados cambios sociales y tecnológicos están impactando fuertemente en la psicometría con la incorporación de la inteligencia artificial, realidad virtual y necesidad de informatizar los instrumentos de medida. Este escenario obliga a replantearse las actuales prácticas en el uso de test y reflexionar en torno a los mecanismos de regulación que deberán implementarse.

## 7. Conclusiones

Las buenas prácticas en el uso de test mínimamente requieren de que el/la psicólogo/a cuente con una sólida formación en el área, que la calidad de los instrumentos esté garantizada y que se utilicen de manera ética y rigurosa. El presente documento abordó una breve revisión de las principales directrices internacionales vinculadas a los tres pilares que coincidentemente diversos autores han identificado como aspectos esenciales para garantizar el buen uso de test psicológicos (Elosua, 2017; Evers et al., 2012; Hernández et al., 2022; Muñiz et al., 2011; Muñiz et al., 2020) y los elementos que a partir de ellos se deberían considerar en el diseño e implementación de una política que promueva las buenas prácticas en el uso de test en Chile.

Primero se resaltó la importancia de que la administración de los test debe estar a cargo de profesionales competentes y calificados/as en cuanto a sus habilidades cognitivas, procedimentales y clínicas, enfatizando la necesidad de fortalecer la formación en el área tanto de pregrado como de postgrado. Este aspecto, por una parte, invita principalmente a que las universidades y otras instituciones se coordinen, establezcan acuerdos y verifiquen su cumplimiento, para ofrecer formación especializada y de calidad en el área. Por otra, resalta la importancia de que el/la psicólogo/a constantemente esté evaluando sus posibilidades y limitaciones, busque fortalecer sus competencias y actualizar sus conocimientos de manera permanente.

En segundo lugar, se describieron algunas de las propiedades psicométricas (evidencias validez, confiabilidad e imparcialidad) que deberían estar garantizadas y que informan sobre la calidad de los test como instrumentos fiables de medida. Esta tarea está principalmente a cargo de los equipos de investigación en el área, ya sea de universidades, laboratorios, centros o sociedades científicas. Sin embargo, esto también es responsabilidad de el/la psicólogo/a que usa los test, puesto que al seleccionar los instrumentos que utilizará en un proceso de evaluación psicológica debe ser capaz de recopilar y evaluar la evidencia psicométrica disponible, y así decidir si le son de utilidad para fundamentar sus inferencias.

Finalmente, se desarrollaron los elementos que mínimamente representan el uso ético y riguroso de test, aludiendo a los aspectos esenciales que se debieran considerar en la adecuada selección y administración de instrumentos, la interpretación y comunicación de sus resultados. Estos aspectos se fundamentan en los dos primeros puntos revisados, dependen esencialmente del ejercicio ético de la profesión que realiza el/la psicólogo/a, pero también requiere de un contexto que permita instalar adecuados mecanismos que regulen el ejercicio profesional.

Un ejemplo de la articulación entre los temas que se han abordado en este trabajo se aprecia en los planteamientos de Juliá (2023), en el ámbito de la Psicología Educacional. La autora enfatiza la necesidad de incorporar la consideración del contexto para lograr generar intervenciones que sean culturalmente pertinentes y situadas en este ámbito del ejercicio profesional. Dentro de esto, señala que los procesos de evaluación psicológica en contextos educacionales deberían ser capaces de respetar la diversidad, garantizar la equidad y salvaguardar que no generen discriminación o segregación. Para la autora es una obligación ética la consideración de los derechos de las personas evaluadas declarados en los estándares internacionales, durante todo el proceso de evaluación, ya que éste fundamenta una “cadena de decisiones” educativas y familiares que impactan la trayectoria educativa de los niños, niñas y adolescentes. Una política nacional que aborde y regule el uso de test dentro de los procesos de evaluación psicológica contribuiría a dar cumplimiento estos requerimientos que plantea la autora, de manera ética, responsable y rigurosa.

Se espera que las ideas desarrolladas en este texto iluminen e impulsen el comienzo de la creación de una política de buenas prácticas en el uso de test en Chile. Esta tarea requiere la articulación de los esfuerzos y el compromiso ético de diversas personas e instituciones, que sin duda aportarán al enriquecimiento del ejercicio profesional de la psicología, pero que sobre resguardarán el derecho de todas las personas a una evaluación justa y pertinente.

## References

- American Education Research Association, American Psychological Association, & National Council on Measurement in Education. (2018). *Estándares para pruebas educativas y psicológicas* (M. Lieve, Trad.). American Educational Research Association.
- American Psychological Association. (2020). *APA Guidelines for Psychological Assessment and Evaluation*. <https://www.apa.org/about/policy/guidelines-psychological-assessment-evaluation.pdf>
- Ato, M., López-García, J. J., & Benavente, A. (2013). Un sistema de clasificación de los diseños de investigación en psicología. *Anales de Psicología*, 29(3), 1038-1059. <https://doi.org/10.6018/analesps.29.3.178511>
- Brenlla, M. E., Seivane, M. S., Fernández Da Lama, R. G., & Germano, G. (2023). Pasos fundamentales para realizar adaptaciones de pruebas psicológicas. *Revista de Psicología*, 19(38), 121-148. <https://doi.org/10.46553/RPSL.19.38.2023.p121-148>
- Carrillo, C. (03 Septiembre 2024). Buscan que todas las profesiones estén colegiadas para combatir faltas a la ética con inhabilidad. *biobiochile.cl*. <https://www.biobiochile.cl/noticias/nacional/chile/2024/09/03/buscan-que-todas-las-profesiones-estén-colegiadas-para-combatir-faltas-a-la-etica-con-inhabilidad.shtml>
- Cataldo, F., Herrera, F., Rojas, D., Rojo, D., Gutiérrez, X., & Bargsted, M. (2012). La evaluación de la propensión al riesgo ¿Es confiable el uso del “test de alerta”? Estudio exploratorio en trabajadores de servicios a la minería. *Salud & Sociedad*, 3(1), 50-64. <https://doi.org/10.22199/S07187475.2012.0001.00004>
- Cortés, J., & Benavente, M. (2007). *Manual de psicodiagnóstico y psicoterapia infantil* (1.ª ed.). RIL Editores.
- Díaz-López, C., Caso, J., & Contreras, L. (2017). Estándares y directrices que orientan el desarrollo y adaptación de instrumentos de medición. In J. C. Rodríguez & J. Caso (Coords.), *Prácticas de investigación aplicada a contextos educativos* (pp. 14-30). Editorial Universitaria.
- Didier, N. (2014). Selección de personal en Chile. *Revista Iberoamericana de Psicología: Ciencia y Tecnología*, 7(1), 103-113. <https://reviberopsicologia.ibero.edu.co/article/view/rip.7109>
- Echavarría-Ramírez, L. M., & Tirapu-Ustárriz, J. (2021). Exploración neuropsicológica en niños con discapacidad intelectual. *Revista de Neurología*, 73(02), 66-76. <https://doi.org/10.33588/rn.7302.2021025>
- Elosua, P. (2003). Sobre la validez de los tests. *Psicothema*, 15(2), 315-321. <https://www.psicothema.com/pdf/1063.pdf>
- Elosua, P. (2012). Test publicados en España: Usos, costumbres y asignaturas pendientes. *Papeles del Psicólogo*, 33(1), 12-21. <https://www.psychologistpapers.com/pdf/2031.pdf>

- Elosua, P. (2017). Avances, proyectos y retos internacionales ligados al uso de tests en Psicología. *Estudios de Psicología (Campinas)*, 34(2), 201-210. <https://doi.org/10.1590/1982-02752017000200002>
- Evers, A., Muñiz, J., Bartram, D., Boben, D., Egeland, J., Fernández-Hermida, J. R., Frans, Ö., Gintiliené, G., Hagemester, C., Halama, P., Iliescu, D., Jaworowska, A., Jiménez, P., Manthouli, M., Matesic, K., Schittekatte, M., Sümer, H. C., & Urbánek, T. (2012). Testing Practices in the 21st Century: Developments and European Psychologists' Opinions. *European Psychologist*, 17(4), 300-319. <https://doi.org/10.1027/1016-9040/a000102>
- Fernández-Ballesteros, R., De Bruyn, E. E. J., Godoy, A., Homke, L. F., Ter Laak, J., Vizcarro, C., Westhoff, K., Westmayer, H., & Zaccagnini, J. L. (2003). Guías para el proceso de evaluación (GAP): Una propuesta a discusión. *Papeles del Psicólogo*, 23(84), 58-70. <https://www.papelesdelpsicologo.es/resumen?pii=1056>
- Fernández-Ballesteros, R., Oliva, M., Vizcarro, C., & Zamarrón, M. (2011). *Buenas prácticas y competencias en evaluación psicológica. El Sistema Interactivo Multimedia de Aprendizaje del Proceso de Evaluación (SIMAPE)*. Pirámide.
- Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., & Muñiz, J. (2022). Decálogo para el Análisis Factorial de los Ítems de un Test. *Psicothema*, 34(1), 7-17. <https://doi.org/10.7334/psicothema2021.456>
- Forns, M., & Amador, J. A. (2017). *Habilidades clínicas para aplicar, corregir e interpretar las escalas de inteligencia de Wechsler*. Pirámide.
- Gregory, R. J. (2012). *Pruebas psicológicas. Historia, principios y aplicaciones* (6.ª Edición). Pearson.
- Guirao-Goris, J., Olmedo, A., & Ferrer, E. (2008). El artículo de revisión. *Revista Iberoamericana de Enfermería Comunitaria*, 1(1), 1-25. [https://www.uv.es/joguigo/valencia/Recerca\\_files/el\\_articulo\\_de\\_revision.pdf](https://www.uv.es/joguigo/valencia/Recerca_files/el_articulo_de_revision.pdf)
- Hall, J. D., Howerton, D. L., & Bolin, A. U. (2005). The Use of Testing Technicians: Critical Issues for Professional Psychology. *International Journal of Testing*, 5(4), 357-375. [https://doi.org/10.1207/s15327574ijto504\\_2](https://doi.org/10.1207/s15327574ijto504_2)
- Hernández, A., Elosua, P., Fernández-Hermida, J. R., & Muñiz, J. (2022). Comisión de Test: Veinticinco años velando por la calidad de los test. *Papeles del Psicólogo*, 43(1), 55-62. <https://doi.org/10.23923/pap.psicol.2978>
- Hernández, A., Hidalgo, M. D., Muñiz, J., & Fernández-Hermida, J. R. (2021). Diferencias de opinión sobre el uso de los test: Su relación con la edad el género y el sector laboral. *Papeles del Psicólogo*, 42(2), 93-104. <https://doi.org/10.23923/pap.psicol.2960>
- Iliescu, D., Bartram, D., Zeinoun, P., Ziegler, M., Elosua, P., Sireci, S., Geisinger, K. F., Odendaal, A., Oliveri, M. E., Twing, J., & Camara, W. (2024). The Test Adaptation Reporting Standards (TARES): Reporting test adaptations. *International Journal of Testing*, 24(1), 80-102. <https://doi.org/10.1080/15305058.2023.2294266>
- International Test Commission. (2013). *International Test Commission Guidelines on Test Use*. [https://www.intestcom.org/files/guideline\\_test\\_use.pdf](https://www.intestcom.org/files/guideline_test_use.pdf)
- Juliá, Z. M. T. (2023). Acción profesional en Psicología Educativa desde la Declaración Universal de Principios Éticos para Psicólogas y Psicólogos. *Revista Liminales. Escritos sobre Psicología y Sociedad*, 12(24), 39-60. <https://doi.org/10.54255/lim.vol12.num24.766>
- Leong, F. T. L., Bartram, D., Cheung, F., Geisinger, K. F., & Iliescu, D. (2020). *Manual internacional de pruebas y evaluación del ITC*. Manual Moderno.
- Loayza, E. F. (2021). El fichaje de investigación como estrategia para la formación de competencias investigativas. *Educare et Comunicare*, 9(1), 67-77. <https://doi.org/10.35383/educare.v9i1.594>
- Manterola, C., Grande, L., Otzen, T., García, N., Salazar, P., & Quiroz, G. (2018). Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. *Revista Chilena de Infectología*, 35(6), 680-688. <https://doi.org/10.4067/S0716-10182018000600680>
- Marín, C. (2021). *Guía práctica de evaluación psicológica clínica. Desarrollo de competencias*. Pirámide.
- Mena, B. (2019). Validez del examen único nacional de conocimientos de Medicina (EUNACOM). In J. Manzi, M. R. García & S. Taut (Eds.), *Validez de las evaluaciones educativas en Chile y Latinoamérica*. Ediciones UC.
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Muñiz, J., & Hambleton, R. K. (1996). Directrices para la traducción y adaptación de los tests. *Papeles del Psicólogo*, 66. <https://www.papelesdelpsicologo.es/resumen?pii=737>
- Muñiz, J., Fernández-Hermida, J. R., Fonseca-Pedrero, E., Campillo-Álvarez, Á., & Peña-Suárez, E. (2011). Evaluación de test editados en España. *Papeles del Psicólogo*, 32(2), 113-128. <https://www.papelesdelpsicologo.es/pdf/1947.pdf>
- Muñiz, J., Hernández, A., Elosua, P., & Wechsler, S. M. (2020). El uso de los Test en España y Latinoamérica: Perspectivas actuales y retos futuros. *Anuario Internacional de Revisiones en Psicología*, 1, 67-94. <https://doi.org/10.14635/REVPSY.0.4>
- Muñiz, J., Hernández, A., & Ponsoda, V. (2015). Nuevas Directrices sobre el Uso de los Test: Investigación, Control de Calidad y Seguridad. *Papeles del Psicólogo*, 36(3), 161-173. <http://www.papelesdelpsicologo.es/pdf/2609.pdf>
- Pardal-Refoyo, J. L. (2023). Los artículos de revisión. Orientaciones para los autores y revisores. *Revista ORL*, 14(3), 171-174. <https://doi.org/10.14201/orl.31646>
- Prieto, G., & Muñiz, J. (2000). Un modelo para evaluar la calidad de los tests utilizados en España. *Papeles del Psicólogo*, 77, 65-75. <https://www.papelesdelpsicologo.es/pdf/2775.pdf>
- Prieto, G., Muñiz, J., Almeida, L. S., & Bartram, D. (1999). Uso de los tests psicológicos en España, Portugal e Iberoamérica. *Revista Iberoamericana de Diagnóstico y Evaluación - e Avaliação Psicológica*, 8(2), 67-82. [https://www.aidep.org/o3\\_riddep/Ro8/Ro84.pdf](https://www.aidep.org/o3_riddep/Ro8/Ro84.pdf)
- Reyes, H. (2020). Artículos de revisión. *Revista Médica de Chile*, 148(1), 103-108. <http://doi.org/10.4067/S0034-98872020000100103>



- Rodríguez, M. J., Navarrete, R., & Bargsted, M. (2017). Problemas éticos y consecuencias reconocidas por psicólogos noveles en reclutamiento y selección de personal. *Psicoperspectivas*, 16(3), 164-176. <https://doi.org/10.5027/psicoperspectivas-vol16-issue3-fulltext-1082>
- Rodríguez-Cancino, M., Vidal-Rivera, C., Navarro-Ovando, V., & Grez-Gaete, O. (2024). *Guía Práctica. Consejos para la aplicación e interpretación de WISC-V*. Pontificia Universidad Católica de Chile. <https://www.cedeti.cl/wp-content/uploads/2024/03/Gui%CC%81a-pra%CC%81ctica-WISC-V-digital-febrero.pdf>
- Sattler, J. M. (2010). *Evaluación Infantil. Fundamentos Cognitivos: Vol. I*. Manual Moderno.
- Sireci, S. G. (2020). De-“Constructing” Test Validation. *Chinese/English Journal of Educational Measurement and Evaluation* | 教育测量与评估双语期刊, 1(1), 1-12. <https://doi.org/10.59863/CKHH8837>
- Sireci, S., & Benítez, I. (2023). Evidence for Test Validation: A Guide for Practitioners. *Psicothema*, 3(35), 217-226. <https://doi.org/10.7334/psicothema2022.477>
- Urbina, S. (2014). *Essentials of Psychological Testing* (2nd Edition). John Wiley & Sons, Inc.
- Vera, O. (2009). Cómo escribir artículos de revisión. *Revista Médica La Paz*, 15(1), 63-69. [http://www.scielo.org.bo/scielo.php?script=sci\\_arttext&pid=S1726-89582011000200010](http://www.scielo.org.bo/scielo.php?script=sci_arttext&pid=S1726-89582011000200010)
- Vinet, E. V., & González, N. (2013). Desarrollos actuales y desafíos futuros en la evaluación psicológica en Chile. *PSIENCIA: Revista Latinoamericana de Ciencia Psicológica*, 5(2), 134-138. <https://psykebase.es/servlet/articulo?codigo=4821189>
- Vinet, E. V., & Rodríguez-Cancino, M. (2024). Evaluación Psicológica en Chile: Historia, Actualidad y Desafíos. In S. Pugliese, J. Livia, & S. Wechsler (Eds.), *Evaluación Psicológica Orígenes y Desafíos en Países Iberoamericanos*. Universidad Ricardo Palma. Editorial Universitaria. Sello Editorial Sociedad Interamericana de Psicología. <https://doi.org/10.30849/SIP.GTEPorigdesibero2024>
- Vinet, E. V., Rodríguez-Cancino, M., Sandoval, A., Rojas-Mora, P., & Saiz, J. L. (2023). El Empleo de Test por Psicólogos/as Chilenos/as: Un Inquietante Panorama. *Psykhe*, 32(1), 1-19. <https://ojs.uc.cl/index.php/psykhe/article/view/25293>
- Wechsler, S. M., Hutz, C. S., & Primi, R. (2019). O desenvolvimento da avaliação psicológica no Brasil: Avanços históricos e desafios. *Avaliação Psicológica*, 18(2). <https://doi.org/10.15689/ap.2019.1802.15466.02>
- Wechsler, S. M., Schultze, M. K., Vinet, E. V., & Livia, J. (2022). Psychological Assessment in South America: Perspectives from Brazil, Bolivia, Chile, and Peru. In S. Laher (Ed.), *International Histories of Psychological Assessment* (1st ed., pp. 295-323). Cambridge University Press. <https://doi.org/10.1017/9781108755078.017>

## Statements

**Funding:** The development of this article is linked to research by the author that was financed by the Dirección de Investigación, Universidad de La Frontera, Project DIUFRO DI22-0030.

**Acknowledgments:** The author is grateful for the guidance and suggestions of Professor José Luis Saiz, Department of Psychology, Universidad de La Frontera, who contributed to the creation of Table 2 on sources of validity evidence.

**Conflicts of Interest:** The author declares no conflict of interest. The funders had no role in the design of the study, the writing of the manuscript, or the decision to publish the results.

**Ethics Committee Review Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Artificial Intelligence Statement:** During the preparation of this manuscript, the author used DeepL, version 24.11.21416769, for editing and proofreading purposes in the Abstract. The author has thoroughly reviewed and revised the output and accepts full responsibility for the content of this publication.