Routledge
Taylor & Francis Group

Measurement and Evaluation in Counseling and Development

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uecd20

Responsibilities of Users of Standardized Tests (RUST-4E)

Prepared for the Association for Assessment and Research in Counseling

A. Stephen Lenz, Haley Ault, Richard S. Balkin, Casey Barrio Minton, Bradley T. Erford, Danica G. Hays, Bryan S. K. Kim & Chi Li

To cite this article: A. Stephen Lenz, Haley Ault, Richard S. Balkin, Casey Barrio Minton, Bradley T. Erford, Danica G. Hays, Bryan S. K. Kim & Chi Li (2022) Responsibilities of Users of Standardized Tests (RUST-4E), Measurement and Evaluation in Counseling and Development, 55:4, 227-235, DOI: 10.1080/07481756.2022.2052321

To link to this article: https://doi.org/10.1080/07481756.2022.2052321

Published online: 15 Apr 2022.

MEASUREMENT AND EVALUATION IN COUNSELING AND DEVELOPMENT
2022, VOL. 55, NO. 4, 227–235
https://doi.org/10.1080/07481756.2022.2052321

Routledge
Taylor & Francis Group

APPLICATIONS OF ASSESSMENT

Responsibilities of Users of Standardized Tests (RUST-4E)

Prepared for the Association for Assessment and Research in Counseling

A. Stephen Lenz[a], Haley Ault[b], Richard S. Balkin[c], Casey Barrio Minton[b], Bradley T. Erford[d], Danica G. Hays[e], Bryan S. K. Kim[f], and Chi Li[g]

[a]Texas A&M University-San Antonio, San Antonio, TX, USA; [b]University of Tennessee Knoxville, Knoxville, TN, USA; [c]The University of Mississippi, University, MS, USA; [d]Vanderbilt University, Nashville, TN, USA; [e]University of Nevada Las Vegas, Las Vegas, NV, USA; [f]University of Hawaii at Hilo, Hilo, HI, USA; [g]The University of Memphis, Memphis, TN, USA

ABSTRACT

In April 2021, The Association for Assessment and Research in Counseling Executive Council commissioned a time-referenced task group to revise the Responsibilities of Users of Standardized Tests (RUST) Statement (3rd edition) published by the Association for Assessment in Counseling (AAC) in 2003. The task group developed a work plan to implement a recursive process for document review, gap analysis, commentary representation, bias accounting, and content revision. While some updates to the document were minor, several substantive enhancements to the structure were made including a re-defined scope, definition of terms, and identification of values that are considered integral across all of the responsibilities. In addition the task group identified and included several updates to the document that are consistent with contemporary representations of evidence for test score validity/reliability and emphases on culturally responsive test administration, variance in score representation, training and supervision, use of technology-mediated testing platforms, and differential consequences of testing.

KEYWORDS

Standardized test; responsibilities; counselor training; ethical issues; measurement

Section 1: Orientation to the Responsibilities of Users of Standardized Tests (RUST-4E)

Scope of the Document

Societal events continue to influence the use of tests and assessment in the counseling profession, including evidence-based and outcome-based counseling practices; the use of standardized high-stakes tests in the educational accountability and reform movements; publication of the *Standards for Educational and Psychological Testing* (4th ed.; American Educational Research Association [AERA], American Psychological Association, National Council on Measurement in Education, 2014); continued proliferation of technology-delivered assessment; and continuing legislative initiatives mandating testing in public schools aligned to state standards.

The continuing purpose of the Responsibilities of Users of Standardized Tests, Fourth Edition (RUST-4E) is to promote the accurate, fair, and responsible use of standardized tests by the counseling and education communities. RUST-4E addresses the needs of the members of the American Counseling Association (ACA) and affiliated counseling organizations, as well as educational professionals, such as teachers and administrators. Test developers, policy makers,

CONTACT A. Stephen Lenz stephen.lenz@tamusa.edu Department of Counseling, Health and Kinesiology, Texas A&M University–San Antonio, One University Way, San Antonio, TX 78224, USA.

and the general public will find this statement useful as they encounter and make decisions about testing issues. The principles in RUST-4E apply to the use of testing instruments regardless of delivery methods (e.g., paper/pencil, technology-mediated) or setting (e.g., group, individual).

RUST-4E helps counselors and other stakeholders implement responsible testing practices by providing a concise statement useful in the ethical practice of testing (ACA, 2014). RUST-4E supports counselors' scope of practice including, but not limited to, "psychotherapy, diagnosis, evaluation; administration of assessments, tests and appraisals; referral; and the establishment of counseling plans for the treatment of individuals, couples, groups and families with emotional, mental, addiction and physical disorders" (Kaplan & Kraus, 2018, p 225). RUST-4E is not meant to reach beyond or reinterpret the principles outlined in the *ACA Code of Ethics* (2014) or *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association, National Council on Measurement in Education, 2014) or to formulate a basis for legal action, but is instead intended for use in collaboration with these documents to inform best practice.

Definition of Terms

Stakeholders

Individuals or groups that bear an interest in testing activities, outcomes, and use of scores. Stakeholders may exist across the ecology of test takers to include test takers themselves; family members; educators; educational administrators; educational, social service, and health systems; and community groups and subgroups.

Standardized Testing Procedures

Processes for administering, scoring, and interpreting standardized tests according to a manual that is consistently reflective of the guidelines and strategies depicted by the developers of standardized tests for use across instances. These procedures commonly include reference to the testing setting, environment, protocol delivery, recording activities, interpretation, and representation of scores.

Standardized Tests

Devices or procedures whose structure, format, content, processes, scores, and uses reflect the application of systematic methods for development, sampling, and interpretation of an operationalized aspect of human development or functioning.

Test Reliability Evidence

The degree to which scores on a test can be expected to remain consistent across multiple administrations which is influenced by the degree that items are related to one another and sources of error. Common sources of reliability evidence include estimates based on internal consistency, agreement between multiple raters, test-retest applications, and use of parallel forms.

Test Takers

Individuals or groups who are the intended participants for standardized tests based on standardized testing procedures and possess certain rights and responsibilities.

Test Users

Testing professionals including counselors and related helping professionals who select, administer, score, interpret, and report tests. These professionals may obtain requisite educational training to competently and ethically complete testing activities.

Test Validity Evidence

The degree to which formal theory, psychometric evidence, and practical applications support interpretations of test scores with individuals or groups for their intended purpose. Common sources of validity evidence include test content, response processes, internal structure, relations with other variables, and consequences of testing.

Values Inherent Across Testing Activities and Processes

Although all counselors who use standardized tests are expected to uphold the ethical standards detailed in the ACA (2014) *Code of Ethics*, several values are integral across all responsibilities described in this document.

Access: Test users acknowledge that standardized tests are not available to everyone. They are cognizant that all individuals in an intended population deserve access to relevant standardized tests regardless of race, ethnicity, and gender identity, geographical location, social status, language, or ability status.

Context: Test users are attuned to contextual factors that shape and are shaped by standardized testing. These factors may include developmental issues, cultural considerations for test users, test takers, and stakeholders and/or the group(s) of which they identify, and testing-related conditions such as test setting, timing, purpose, and use.

Efficiency: Test users recognize resource scarcity (e.g., financial and time costs of tests) and communicate test results in a manner that facilitates intended use without undue burden on test takers, test users, or stakeholders. In this sensitivity, they remain flexible in testing activities to allow for optimal use of available resources.

Fairness: Test users act in ways that promote equitable treatment during testing activities while accounting for the influence of measurement bias, differential representation of items, constructs, and consequences of scores within and across groups.

Integration: Test users affirm that the results of standardized tests should be unified into the larger representation of test takers along the developmental continuum.

Proportionality: Test users display a reasonable sense of weight and balance between insights gleaned from scores on standardized measures and other sources of information within the broader assessment process.

Advocacy Orientation: Test users approach testing activities and processes through a lens of empowerment. With an understanding of test taker rights and responsibilities, they effectively communicate to test takers and other stakeholders how standardized testing may be used as a tool for advocacy. Advocacy efforts can include, but are not limited to, promoting individual growth and empowerment, addressing inequities, and raising awareness of cultural ways of knowing and expression of characteristics evaluated through testing.

Commitment to Cultural Competency: Test users value their ongoing learning regarding individual and contextual considerations to ensure culturally-responsive testing activities and processes. They remain well-versed in relevant scholarship and practices related to standardized testing, possess humility regarding their cultural competency, and strive to apply *Multicultural and Social Justice Counseling Competencies* (Ratts et al., 2016) throughout the standardized testing process.

Social Significance and Implications of Responsible Use of Standardized Tests

Test users have foundational coursework and qualifications to responsibly engage in testing activities. As such, test users are cognizant of historical and contemporary developments and use trends by testing category as well as the overall testing process. In addition, they attend to ethical, legal, professional, and multicultural considerations in testing activities. Some of these considerations include test access, professional licensure, technology use and accessibility, test

fairness, and cultural bias. Furthermore, state statutes determine test users' scope of practice, potentially limiting test users' ability to engage in testing and other assessment practices consistent with their training. All stakeholders have the potential to benefit from ongoing professional legislative and social advocacy to preserve or expand test users' scope of practice regarding testing activities.

Responsible use of standardized tests ensures appropriateness use of results which may influence stakeholders in a variety of ways. Most immediately, tests may provide information regarding test taker strengths, interests, diagnoses, and areas in need of clinical attention. In many settings, results are used to influence access to services, make educational decisions, and inform professional practice. Tests may also inform understanding of program impact and effectiveness and influence resource allocation within accountability mechanisms.

Section 2: Activities Associated with the Responsible Use of Standardized Tests

Qualifications of Test Users

Test users engage with standardized tests for which they have been trained and are competent to select, administer, score, interpret, and/or communicate their findings. If necessary, test users obtain permission from publishers before administering tests. They adhere to the highest degree of ethical codes, laws, and standards governing professional practice. As such, test users participate in ongoing assessment training, cultural competency development, and supervision and/or consultation.

Qualified test users advocate for the well-being of clients and understand that testing is only one component of professional practice. Test users carefully weigh the potential social, cultural, and other contextual consequences of a test for a particular test taker or group of test takers. Test users inform test takers of the benefits and challenges of a particular test and respect test takers' decisions or choices.

In addition to general assessment competencies, test users are familiar with test-specific characteristics, including psychometric, cultural, and other contextual strengths and limitations. They communicate this information to test takers as appropriate throughout the testing process.

Technical Knowledge

Responsible use of tests requires technical knowledge obtained through training, education, and continuing professional development. Test users must seek consultation and/or supervision when interpreting test data outside their scope of training. Responsible test users are conversant and competent in technical aspects of testing including:

- **Validity of Test Results:** Validity is the integrated accumulation of evidence to support a specific interpretation of test results and use of the results. Because validity is a characteristic of test results, a test may have validity evidence of varying degrees, for different purposes. The concept of instructional validity relates to how well the test is aligned to state standards and classroom instructional objectives. Test users should be aware of current standards for evaluating validity evidence (i.e., evidence of test content, response processes, internal structure, relationships to other test scores, and consequences of test use).
- **Reliability:** Reliability refers to the accuracy and consistency of test scores. Various methods are used to calculate and estimate reliability depending on the purpose for which the test is used. Reliability is a function of scores, not scale. Hence, when communicating information about reliability, estimates should be addressed in relation to scores from a normative sample or research sample. Reliability estimates may vary across samples and responsible test use includes the practical application of reliability as a function of scores,

not the tests themselves. Therefore, particular attention is warranted determining and describing test score reliability in relation to test takers who may not have been sufficiently represented in the normative sample or extant research.

- **Confidence Intervals:** Due to the nature of human interaction with standardized tests, a "true" score may never be known. Error may be caused by test characteristics, flawed scoring measures, environment, or intrapersonal factors of the test user and/or taker. Confidence intervals, derived from errors of measurement, describe how an individual's observed score may be distributed over repeated administrations to provide a more informed interpretation of test scores when making decisions impacting access to education, treatment, and resources. Test users should address interpretation of confidence intervals when communicating results, especially when confidence intervals provide a range of classifications or potential diagnostic considerations.
- **Scores and Norms:** When using norm-referenced instruments, test users determine the appropriateness of the norming group based on client factors, purpose of assessment, and use of scores. Additional factors contributing to norm adequacy are sampling method, sample size, and demographic characteristics in relation to age, gender, ethnicity, educational background, socioeconomic status, among others.

Test Selection

Responsible test selection requires establishing a clear and defensible connection between (a) the nature and purpose of a testing request, (b) user qualifications, (c) expected characteristics of the testing environment, (d) test content, processes, and intended use of test scores, and (e) test taker characteristics. Once a testing request has been received, test users should clarify the broader context surrounding the nature of a testing request and verify the specific purpose for which the testing activity is intended. Common purposes for a testing request include:

- **Description:** Obtaining objective information on the status of certain characteristics such as achievement, ability, psychological symptoms, career, personality, and system-level characteristics;
- **Prediction:** Supporting determinations of eligibility for certain programs, eligibility for educational and vocational opportunities and interventions, access to institutional resources, and other outcomes of interest;
- **Accountability:** Judging the progress of an individual or the effectiveness of an educational institution. Strong alignment between what is targeted and what is tested needs to be present; likewise, strong alignment between condition and instrumentation is needed when determining the effectiveness of interventions.
- **Program Planning and Evaluation:** Supporting program planning and evaluation activities in conjunction with other data sources. The role of testing in collection of formative and summative data is key for assuring program quality, opportunities for improvement, and estimations of impact.

With the context and purpose clarified, test users are encouraged to base test selection on a number of necessary, but not singularly sufficient, factors. Primary factors include qualifications addressed in expected characteristics of the testing environment and intended use of test scores (see Section 2.I). Responsible test use involves identification of expected characteristics of the testing environment and selection of a test that can be administered with a degree of fidelity and consistency that promotes the representativeness of a test taker's scores and the underlying constructs. With environmental considerations accounted for, test review and selection involves determining the degree to which the content, processes, and intended uses are appropriate for the intended audience and are of sufficient technical quality for the purpose at hand. This may

involve critical review of instructions, items, and standardized procedures to assure contemporary terminology, developmentally-appropriate tasks, and culturally-representative language and scenarios. Any limitations based on these considerations should be identified and described in the testing report. Some areas to consider include the following:

- **Test Taker Characteristics:** Test users review technical information to determine if the test characteristics are appropriate for the test taker. Considerations include age, grade level, ability status, language proficiency, reading level, and cultural identity;
- **Well-Defined Procedures for Administration:** Test users only select tests that use well-defined scoring procedures that promote accuracy and consistency across administrations. This includes ensuring that translated and cross-cultural adaptations bear sufficient evidence for equivalence to the identified test taking scenario. When using technology-mediated scoring services, test users review and verify procedures to provide adequate clarity for use in the testing setting;
- **Normative Group Representativeness:** Test users review the normative group's representativeness to determine the degree of appropriateness for interpreting the intended test takers' scores. When a test taker is outside of or minimally represented within the normative sample, test users evaluate additional evidence to determine appropriateness of test use;
- **Modifications:** When test takers have health conditions or disabilities, test users may find alternative measures and/or employ accommodations in test taking procedures.

In instances when a test user is providing services under supervision, it is prudent to verify the connection between test purpose and selection activities with their supervisor. Additionally, when an autonomous or independent test user is selecting a test for an emerging, novel, or proxy use, seeking consultation or supervision from an individual or group that is proficient in administration, scoring, interpretation, and reporting procedures is recommended.

Test Administration

Responsible test administration includes carefully following standardized procedures and instructions described by the test developers to align test use practices with intended practices upon which validity and reliability data are based. Test users should ensure that test takers work within conditions that maximize opportunity for optimum performance which includes awareness, knowledge, and use of developmentally-informed, ability-referenced, and culturally-responsive test administration practices. As appropriate, stakeholders should be involved in various aspects of the testing process that include the following:

Before administration, it is important that relevant persons:

- Are informed about the standard testing procedures, including information about the purposes of the test, the kinds of tasks involved, and the methods of administration, scoring, and reporting. Informed consent should also include the purpose and timing of testing, uses of scores, and discussion of transfer for automated scoring services if relevant;
- Have sufficient practice experiences prior to the test to include preparation, as needed, on how to operate equipment for computer-administered tests and rehearsal in responding to tasks;
- Demonstrate sufficient training in test user responsibilities and administration procedures for the test;

- Review test materials and procedures within the context of administration sites prior to test use to ensure standardized conditions and plan for responses that minimize influence on score validity in the event that any irregularities occur;
- Receive and prepare for appropriate accommodations or modifications of testing materials and procedures in order to accommodate test takers with disabilities; and
- Communicate a clear understanding of test taker rights and responsibilities.

During administration, it is important that

- The testing environment (e.g., seating, work surfaces, lighting, room temperature, freedom from distractions) and psychological status are conducive to the best possible performance of the examinees;
- Sufficiently trained personnel establish and maintain uniform conditions and observe the conduct of test takers when large groups of individuals are tested;
- Test users follow the instructions in the test manual, demonstrate verbal clarity, use verbatim directions, adhere to verbatim directions, follow exact sequence and timing, and use materials that are identical to those specified by the test publisher. Any variation from these practices should be documented and accounted for during scoring, interpretation, and reporting activities to minimize influence of variations of scores associated with construct-irrelevant responding;
- A systematic and objective procedure is in place for observing and recording environmental, health, and emotional factors, or other elements that may invalidate test performance and results; deviations from prescribed test administration procedures, including information on test accommodations for individuals with disabilities, are recorded;
- The security of test materials and computer-administered testing software is protected, ensuring that only individuals with a legitimate need for access to the materials/software can obtain such access; steps to eliminate the possibility of breaches in test security and copyright protection are respected.

After administration, it is important to

- Collect and inventory all secure test materials and immediately report any breaches in test security;
- Include notes on any problems, irregularities, and accommodations in the test records;
- Inform test users about the consistency of the test administration with its intended procedures and practices;
- Provide description about the timeline, format, and general content of test score reporting and communication activities. These precepts represent the basic process for all standardized tests and assessments. Some situations may add steps or modify some of these to provide the best testing milieu possible.

Test Scoring

Tests may incorporate different scoring procedures (e.g., automated/computer scoring, hand scoring, third-party scoring, and self-scoring), and each procedure has advantages and disadvantages. Test takers should be informed of different scoring options before testing, and engage in dialogue regarding which approach best fits their needs. In some cases, such as symptom monitoring inventories, test taker inclusion in the scoring procedure may have educational or therapeutic value. However, in other cases, the nature of test qualifications and technical proficiency requirements may preclude test taker inclusion in the scoring activities. Regardless, test administrators should be aware of how personal values and biases may influence this decision-making process and engage in strategies to mitigate related influence. Test users should

follow the scoring procedures published by the test developers to ensure scoring accuracy and consistency. This extends to use of computer-based scoring services when incumbent upon the test user to review and verify the consistency of raw data entered with report materials.

Interpreting Test Results

Responsible test interpretation requires knowledge about and experience with the test, scores, decisions to be made, and test taker. Interpretation of scores on any test should be made alongside a thorough knowledge of the technical aspects of the test, test results, and limitations. Many factors can impact the valid and useful interpretations of test scores. These can be grouped into several categories including psychometric, test taker, and contextual.

- **Psychometric Factors:** Factors such as the reliability, norms, standard error of measurement, and validity are important when interpreting test results. This includes careful attention to measurement invariance. Test users interpret scores with caution for test takers who are underrepresented in norming groups. Responsible test users consider these basic concepts and how each impacts scores and hence interpretation of test results.
- **Test Taker Factors:** Factors such as the test taker's social location, cultural identities, experience with testing, and investment in testing may impact results obtained from the test. Specifically, test users should evaluate how the test taker's racial and ethnic identity, gender identity, sexual/affectional orientation, age, social class, ability status, religious/spiritual beliefs, and marital status, impact their results. Responsible test users consider these factors as critical when interpreting test results.
- **Contextual Factors:** Test users should integrate influence of accommodations or omitted scales on scores and interpretation. The relationship of the test to the instructional program, opportunity to learn, quality of the educational program, work and home environment, and other factors that would assist in understanding the test results are useful in interpreting test results. For example, if the test does not align to curriculum standards and how those standards are taught in the classroom, the test results may not provide useful information.

The process of interpretation is meant to be synergistic. Quality interpretation includes triangulation of multiple sources of data and collaboration with stakeholders.

Communicating Test Results

Communicating test results is an interactive, multi-modal, and culturally-responsive process between the test taker, test user, and other stakeholders as relevant. Test users present test findings with relevant contextual information, using multiple strategies (e.g. verbal, visual) involving the test taker and/or other stakeholders when deciding how the findings may be integrated with other client information. The goal of communicating test results is to share information that can be used to foster client self-awareness, interventions aligned to client goals, stakeholder knowledge, and client and social advocacy.

Prior to communicating test results, the test user is well-versed in general test interpretation principles as well as an understanding of how to interpret the particular test being used. In addition, test users are familiar with the most updated scholarship and best practices regarding integrating a particular test's findings and communicating those findings in an empowering and accessible manner. As such, the test taker is invited into a conversation about the test findings to ensure client understanding, offer pertinent individual and other contextual information, and co-generate ideas for test use.

Disclosure Statement

No potential conflict of interest was reported by the authors.

Notes on Contributors

A. Stephen Lenz is Chair and Professor in the Department of Counseling, Health and Kinesiology at Texas A&M University-San Antonio.

Haley Ault is a Doctoral Candidate in the Educational Psychology and Counseling Department at the University of Tennessee, Knoxville.

Richard S. Balkin is Chair and Professor in the Department of Leadership and Counselor Education at The University of Mississippi.

Casey Barrio Minton is a Professor in the Educational Psychology and Counseling Department at the University of Tennessee, Knoxville.

Bradley T. Erford is Program Director and Professor in the Department of Human and Organizational Development at Vanderbilt University.

Danica G. Hays is Dean and Professor in the College of Education at the University of Nevada, Las Vegas.

Bryan S. K. Kim is Chair and Professor in the Division of Social Sciences at the University of Hawaii at Hilo.

Chi Li is an Assistant Professor in the Department of Counseling, Educational Psychology and Research at The University of Memphis.

ORCID

A. Stephen Lenz http://orcid.org/0000-0002-3360-835X
Richard S. Balkin http://orcid.org/0000-0003-4134-6385
Bradley T. Erford http://orcid.org/0000-0001-5891-5770
Danica G. Hays http://orcid.org/0000-0003-3685-5177
Chi Li http://orcid.org/0000-0002-1128-0883

References

American Counseling Association (ACA). (2014). *ACA code of ethics*. Author.

American Educational Research Association [AERA], American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. AERA.

Kaplan, D. M., & Kraus, K. L. (2018). Building blocks to portability: Culmination of the 20/20 initiative. *Journal of Counseling & Development*, *96*(2), 223–228. https://doi.org/10.1002/jcad.12195

Ratts, M. J., Singh, A. A., Nassar-McMillan, S., Butler, S. K., & McCullough, J. R. (2016). Multicultural and social justice counseling competencies: Guidelines for the counseling profession. *Journal of Multicultural Counseling and Development*, *44*(1), 28–48. https://doi.org/10.1002/jmcd.12035