

Assessing The Reliability Validity And Use Of The

The internal validity of a study reflects the extent to which the design and conduct of the study have prevented bias(es). One of the key steps in a systematic review is assessment of a study's internal validity, or potential for bias. This assessment serves to: (1) identify the strengths and limitations of the included studies; (2) investigate, and potentially explain heterogeneity in findings across different studies included in a systematic review; and (3) grade the strength of evidence for a given question. The risk of bias assessment directly informs one of four key domains considered when assessing the strength of evidence. With the increase in the number of published systematic reviews and development of systematic review methodology over the past 15 years, close attention has been paid to the methods for assessing internal validity. Until recently this has been referred to as "quality assessment" or "assessment of methodological quality." In this context "quality" refers to "the confidence that the trial design, conduct, and analysis has minimized or avoided biases in its treatment comparisons." To facilitate the assessment of methodological quality, a plethora of tools has emerged. Some of these tools were developed for specific study designs (e.g., randomized controlled trials (RCTs), cohort studies, case-control studies), while others were intended to be applied to a range of designs. The tools often incorporate characteristics that may be associated with bias; however, many tools also contain elements related to reporting (e.g., was the study population described) and design (e.g., was a sample size calculation performed) that are not related to bias. The Cochrane Collaboration recently developed a tool to assess the potential risk of bias in RCTs. The Risk of Bias (ROB) tool was developed to address some of the shortcomings of existing quality assessment instruments, including over-reliance on reporting rather than methods. Several systematic reviews have catalogued and critiqued the numerous tools available to assess methodological quality, or risk of bias of primary studies. In summary, few existing tools have undergone extensive inter-rater reliability or validity testing. Moreover, the focus of much of the tool development or testing that has been done has been on criterion or face validity. Therefore it is unknown whether, or to what extent, the summary assessments based on these tools differentiate between studies with biased and unbiased results (i.e., studies that may over- or underestimate treatment effects). There is a clear need for inter-rater reliability testing of different tools in order to enhance consistency in their application and interpretation across different systematic reviews. Further, validity testing is essential to ensure that the tools being used can identify studies with biased results. Finally, there is a need to determine inter-rater reliability and validity in order to support the uptake and use of individual tools that are recommended by the systematic review community, and specifically the ROB tool within the Evidence-based Practice Center (EPC) Program. In this project we focused on two tools that are commonly used in systematic reviews. The Cochrane ROB tool was designed for RCTs and is the instrument recommended by The Cochrane Collaboration for use in systematic reviews of RCTs. The Newcastle-Ottawa Scale is commonly used for nonrandomized studies, specifically cohort and case-control studies.

Improving the accuracy of a survey is the focus of Mark S. Litwin's book, which shows how to assess and interpret the quality of survey data by thoroughly examining the survey instrument used. He explains how to code and pilot test new and established surveys. In addition, he covers issues such as: how to measure reliability (including test-retest, alternate form, internal consistency, inter-observer and intra-observer reliability); how to measure validity (including content, criterion and construct validity); how to address cross-cultural issues in survey research; and how to scale and score a survey.

This is an essential book for all those concerned with the field of assessment. It addresses relevant and timely conceptual and practical issues from a research perspective and, based on research results, clearly provides solutions to practical applications at the cutting edge of the emerging area of new modes of assessment. In a clear and rigorous manner, the authors explore new methods and study the various quality aspects of innovative approaches.

Infant and Preschool Assessment Instruments

Reliability, Validity, and Utility

Reliability, Validity and Application of an Improved Scale for Assessment of Motion Sickness Severity

Assessing Speaking Skills

Issues of Feasibility, Reliability, Validity and Bias

This book investigates how strategic marketing is influencing organizations' innovation performance. It presents a structured study conducted on micro, small and medium enterprises (MSMEs) to identify the correlation between innovation and strategic marketing in both Indian manufacturing and service organizations. Strategic marketing, which achieves excellence in organizations, is at the heart of improving productivity, profitability and market sustainability, while also supplying organizations a much sought-after competitive advantage. It has been observed that strategic marketing can significantly help any manufacturing firm to attain the required level of growth, and can specifically be used by Indian firms as a means of deriving differential advantages. Indian organizations that pursue more traditional approaches to strategy are now feeling the intense market pressure to modify their policy with respect to quality. The result has been a shift to 'Strategy with Innovation'. As market competition has again increased, providing quality products and services alone is no longer sufficient; companies are additionally expected to provide products with a touch of innovation to attract customers. This calls for innovation in terms of products, processes, technologies, systems and practically every other area of an organization, which can only be achieved by combining strategies and innovation. This book identifies the variables of strategic marketing that can help companies significantly boost their innovation performance, making it a valuable resource for students, researchers, professionals and policymakers alike.

PROVEN STRATEGIES FOR REVOLUTIONIZING HEALTHCARE SYSTEMS "If I had to sum up this book in one word, the word would be 'brilliant'! This is one of the most insightful books on TOC, not just for healthcare, that I have ever read." --BOB SPROULL, author of The Ultimate Improvement Cycle: Maximizing Profits through the Integration of Lean, Six Sigma, and the Theory of Constraints Performance Improvement for Healthcare: Leading Change with Lean, Six Sigma, and Constraints Management lays out an integrated approach for using three industrially based methods to transform hospital operations in terms of patient outcomes and

experience, financial viability, and employee satisfaction. This pioneering guide presents a scalable strategy for managing bottlenecks, eliminating waste, reducing errors, and containing costs in healthcare organizations, as well as sustaining the gains achieved. Real-world case studies illustrate successful performance improvement implementations that have realized breakthrough operational and financial results. **COVERAGE INCLUDES:** Constraints Management applications in healthcare The NOVACES SystemCPI--an integrated performance improvement deployment approach Three-part assessment--strategic gap analysis, system-level value stream analysis, and system constraint analysis Planning a performance improvement program deployment to ensure timely and consistent execution Applying the right tool to the right problem from a system perspective Sustaining gains achieved by the performance improvement team Defining a path to self-sufficiency

The Understanding Research series focuses on the process of writing up social research. The series is broken down into three categories: Understanding Statistics, Understanding Measurement, and Understanding Qualitative Research. The books provide researchers with guides to understanding, writing, and evaluating social research. Each volume demonstrates how research should be represented, including how to write up the methodology as well as the research findings. Each volume also reviews how to appropriately evaluate published research. **Validity and Validation** is an introduction to validity theory and to the methods used to obtain evidence for the validity of research and assessment results. The book pulls together the best thinking from educational and psychological research and assessment over the past 50 years. It briefly describes validity theory's roots in the philosophy of science. It highlights the ways these philosophical perspectives influence concepts of internal and external validity in research methodology, as well as concepts of validity and reliability in educational and psychological tests and measurements. Each chapter provides multiple examples (e.g., research designs and examples of output) to help the readers see how validation work is done in practice, from the ways we design research studies to the ways we interpret research results. Of particular importance is the practical focus on validation of scores from tests and other measures. The book also addresses strategies for investigating the validity of inferences we make about examinees using scores from assessments, as well as how to investigate score uses, the value implications of score interpretations, and the social consequences of score use. With this foundation, the book presents strategies for minimizing threats for validity as well as quantitative and qualitative methods for gathering evidence for the validity of scores.

Validity and Reliability of the School Assessment of Motor and Process Skills
Evaluation of Three Occlusal Indices for Reliability, Validity and Applicability
Reliability and Validity in Qualitative Research
Reliability and Validity Assessment

Lecturers, request your electronic inspection copy to review it for your course. Validity is the hallmark of quality for educational and psychological measurement. But what does quality mean in this context? And to what, exactly, does the concept of validity apply? These apparently innocuous questions parachute the unwary inquirer into a minefield of tricky ideas. This book guides you through this minefield, investigating how the concept of validity has evolved from the nineteenth century to the present day. Communicating complicated concepts straightforwardly, the authors answer questions like: What does 'validity' mean? What does it mean to 'validate'? How many different kinds of validity are there? When does validation begin and end? Is reliability a part of validity, or distinct from it? This book will be of interest to anyone with a professional or academic interest in evaluating the quality of educational or psychological assessments, measurements and diagnoses.

This open access book offers up-to-date advice and practical guidance on how to undertake a discrete choice experiment as a tool for environmental valuation. It discusses crucial issues in designing, implementing and analysing choice experiments. Compiled by leading experts in the field, the book promotes discrete choice analysis in environmental valuation through a more solid scientific basis for research practice. Instead of providing strict guidelines, the book helps readers avoid common mistakes often found in applied work. It is based on the collective reflections of the scientific network of researchers using discrete choice modelling in the field of environmental valuation (www.envecho.com).
Reliability and Validity of International Large-Scale Assessment
Understanding IEA's Comparative Studies of Student Achievement
Springer Nature

Scale Development

Assessing in VET

Validity and Validation

Validity in Educational and Psychological Assessment

Performance Assessment in Strength and Conditioning

Kirk and Miller define what is -- and what is not -- qualitative research. They suggest that the use of numbers in the process of recording and analyzing observations is less important than that the research should involve sustained interaction with the people being studied, in their own language and on their own turf. Following a chapter on objectivity, the authors discuss the role of reliability and validity and the problems that arise when these issues are neglected. They present a paradigm for the qualitative research process that makes it possible to pursue validity without neglecting reliability.

This open access book describes and reviews the development of the quality control mechanisms and methodologies associated with IEA's extensive program of educational research. A group of renowned international researchers, directly involved in the design and execution of IEA's international large-scale assessments (ILSAs), describe the operational and quality control procedures that are employed to address the challenges associated with providing high-quality, comparable data. Throughout the now considerable history of IEA's international large-scale assessments, establishing the quality of the data has been paramount. Research in the complex multinational context in which IEA studies operate imposes significant burdens and challenges in terms of the methodologies and technologies that have been developed to achieve the stated study goals. The demands of the twin imperatives of validity and reliability must be satisfied in the context of multiple and diverse cultures, languages, orthographies, educational structures, educational histories, and traditions. Readers will learn about IEA's approach to such challenges, and the methods used to ensure that the quality of the data provided to policymakers and researchers can be trusted. An often neglected area of investigation, namely the consequential validity of ILSAs, is also explored, examining issues related to reporting, dissemination, and impact, including discussion of the limits of interpretation. The final chapters address the question of the influence of ILSAs on policy and reform in education, including a case study from Singapore, a country known for its outstanding levels of achievement, but which nevertheless seeks the means of continual improvement, illustrating best practice use of ILSA data.

Assessing the Reliability of Computer-Processed Data (Supersedes GAO-03-273G)

Advancing Human Assessment

The Oxford Handbook of Assessment Policy and Practice in Music Education

Assessing the Reliability and Validity of Multi-Attribute Utility Procedures: An

Application of the Theory of Generalizability

Research Methods in Psychology

Managing Air Quality And Scenic Resources At National Parks And Wilderness Areas

Traditionally, neonates were thought to lack the capacity to perceive pain. Research has since refuted this notion, and there is now an urgent need for improved pain assessment practices, particularly in the neonatal intensive care unit (NICU). While there is an abundance of published neonatal pain assessment measures, few have undergone extensive psychometric testing, and there is a dearth of studies formally assessing their clinical utility. This study aimed to investigate the internal consistency, inter-rater reliability, concurrent and construct validity, and the clinical utility of the COVERS scale and the modified Pain Assessment Tool (PAT) within the NICU of Auckland City Hospital. It also aimed to establish the scales' reliability and validity when used among extremely preterm, preterm, and term neonates. Eighty infants undergoing a routine blood-collection procedure were assessed at baseline and during heel lance by two, independent nurses using the two pain measures. Both scales had satisfactory internal consistency, and high agreement between the two raters. Total scores on each scale correlated highly, and scores correlated with pain ratings provided by mothers and the researcher, which supports both scales' concurrent validity. Support for the construct validity of both scales was gained by results showing that the scales could discriminate effectively between pain and non-pain stimuli. Interestingly, neither scale had acceptable inter-rater reliability or concurrent validity for term infants at baseline, and the concurrent validity of both scales among extremely preterm infants during heel lance was moderate. Relative to the modified PAT, the COVERS scale also demonstrated weaker internal consistency among extremely preterm infants, and weaker inter-rater reliability among preterm infants. Following the study with infants, 25 nurse participants completed a questionnaire about their perceptions of the clinical utility of each scale. Both scales received positive evaluations from nurses. However, the modified PAT was rated as more appropriate to the range of infants admitted to the NICU, and was the preferred tool in this study. This research has implications for the clinical assessment of neonatal pain. The results suggest that a one-size-fits-all approach to neonatal pain assessment may not be feasible. Results also show that both the COVERS scale and the modified PAT represent viable means of assessing neonatal pain in the intensive care unit; however, the modified PAT is more suited to the NICU of Auckland City Hospital. No other book reviews clinical neuropsychological assessment from an empirical psychometric perspective. In this completely revised and updated 2nd edition, the concepts and methods of psychometric neuropsychology are presented as a framework by which to evaluate current instruments. Newer methodologies and statistical techniques are discussed, such as meta analysis, effect size, confirming factor analysis and ecological validity. The explosion of research in this area since the publication of the first edition in 1989, has been incorporated, including a greatly expanded chapter on child assessment instruments. This volume is a must for the bookshelf of every clinical neuropsychologist as well as researchers and students. Anyone conducting forensic evaluations will especially find useful the information on reliability and validity when preparing for court appearances.

A comprehensive textbook for research methods classes. This book is a peer-reviewed inter-institutional project

Assessing the Reliability of Computer-Processed Data (Supersedes Gao-03-273g)

The Methodological, Psychological and Policy Contributions of ETS

Encyclopedia of Behavioral Medicine

Criterion-referenced Test Development

Assessing the Validity and Test-retest Reliability for "pick K of N" Data

BACKGROUND: Numerous tools exist to assess methodological quality, or risk of bias in systematic reviews; however, few have undergone extensive reliability or validity testing.

OBJECTIVES: (1) assess the reliability of the Cochrane Risk of Bias (ROB) tool for randomized controlled trials (RCTs) and the Newcastle-Ottawa Scale (NOS) for cohort studies between individual raters, and between consensus agreements of individual raters for the ROB tool; (2) assess the validity of the Cochrane ROB tool and NOS by examining the association between study quality and treatment effect size (ES); (3) examine the impact of study-level factors on reliability and validity. **METHODS:** Two reviewers independently assessed risk of bias for 154 RCTs. For a subset of 30 RCTs, two reviewers from each of four Evidence-based Practice Centers assessed risk of bias and reached consensus. Inter-rater agreement was assessed using kappa statistics. We assessed the association between ES and risk of bias using meta-regression. We examined the impact of study-level factors on the association between risk of bias and ES using subgroup analyses. Two reviewers independently applied the NOS to 131 cohort studies from 8 meta-analyses. Inter-rater agreement was calculated using kappa statistics. Within each meta-analysis, we generated a ratio of pooled estimates for each quality domain. The ratios were combined to give an overall estimate of differences in effect estimates with inverse-variance weighting and a random effects model. **RESULTS:** Inter-rater reliability between two reviewers was considered fair for most domains (κ ranging from 0.24 to 0.37), except for sequence generation ($\kappa=0.79$, substantial). Inter-rater reliability of consensus assessments across four reviewer pairs was moderate for sequence generation ($\kappa=0.60$), fair for allocation concealment and "other sources of bias" ($\kappa=0.37, 0.27$), and slight for the remaining domains (κ ranging from 0.05 to 0.09). Inter-rater variability was influenced by study-level factors including nature of outcome, nature of intervention, study design, trial hypothesis, and funding source. Inter-rater variability resulted more often from different interpretation of the tool rather than different information identified in the study reports. No statistically significant differences were found in ES when comparing studies categorized as high, unclear or low risk of bias. Inter-rater reliability of the NOS varied from substantial for length of followup to poor for selection of non-exposed cohort and demonstration that the outcome was not present at outset of study. We found no association between individual NOS items or overall NOS score and effect estimates. **CONCLUSION:** More specific guidance is needed to apply risk of bias/quality tools. Study-level factors that were shown to influence agreement provide direction for detailed guidance. Low agreement across pairs of reviewers has implications for incorporation of risk of bias into results and grading the strength of evidence. Variable agreement for the NOS, and lack of evidence that it discriminates studies that may provide biased results, underscores the need for more detailed guidance to apply the tool in systematic reviews.

This guide explains how social scientists can evaluate the reliability and validity of empirical measurements, discussing the three basic types of validity: criterion related, content, and construct. In addition, the paper shows how reliability is assessed by the retest method, alternative-forms procedure, split-halves approach, and internal consistency method.

"Sponsored by the International Testing Commission, The ITC Handbook of International Testing and Assessment is dedicated to the advancement of theory, research, and practice

in the area of international testing and assessment in psychology, education, counseling, organizational behavior, human resource management, and related disciplines"--

Guidance on Design, Implementation and Data Analysis

Strategic Marketing and Innovation for Indian MSMEs

Evaluating the Reliability, Validity, and Clinical Utility of Two Measures of Neonatal Pain

Reliability, Validity and Applications of an Improved Scale for Assessment of Motion

Sickness Severity

A Primer

This report presents a theoretical rationale for assessing the reliability, validity, and dependability of multi-attribute utility models and techniques. If an investigator is advocating the use of a MAU model or procedure he or she is interested in generalizing from observations at hand to a universe or domain of observations that are members of that same universe. The universe must be unambiguously defined but it is not necessary to assume that universe as having any statistical properties such as uniform variance or covariances. A study of generalizability is conducted by taking measurements on persons, stimuli, tasks, etc. that are assumed to be randomly representative of a universe an investigator wishes to generalize to. The ratio of an estimate of the universe 'score' variance to an estimate of the observed score variance is the coefficient of generalizability. This is estimated by the intra-class correlation coefficient. ANOVA and the Expected Mean Square paradigm of Cornfield and Tukey is used to obtain the appropriate variance estimates. The theory dispenses with unnecessary and unwarranted assumptions, and eliminates the distinction between reliability and validity.

In the music classroom, instructors who hope to receive aid are required to provide data on their classroom programs. Due to the lack of reliable, valid large-scale assessments of student achievement in music, however, music educators in schools that accept funds face a considerable challenge in finding a way to measure student learning in their classrooms. From Australia to Taiwan to the Netherlands, music teachers experience similar struggles in the quest for a definitive assessment resource that can be used by both music educators and researchers. In this two-volume Handbook, contributors from across the globe come together to provide an authority on the assessment, measurement, and evaluation of student learning in music. The Handbook's first volume emphasizes international and theoretical

perspectives on music education assessment in the major world regions. This volume also looks at technical aspects of measurement in music, and outlines situations where theoretical foundations can be applied to the development of tests in music. The Handbook's second volume offers a series of practical and US-focused approaches to music education assessment. Chapters address assessment in different types of US classrooms; how to assess specific skills or requirements; and how assessment can be used in tertiary and music teacher education classrooms. Together, both volumes of The Oxford Handbook of Assessment in Music Education pave the way forward for music educators and researchers in the field.

The literature on issues of the reliability and validity of objective and performance assessments in vocational education and training was reviewed and synthesized to develop guidelines for procedures to enhance reliability and validity. The review established the following facts: (1) validity is determined largely through inferences made by both the task developers and users; (2) the assessor's intention is an essential component of validity; (3) although "authenticity" provides evidence of the face validity of workplace assessments, additional evidence of content, criterion, construct, and consequential validity is needed before assessments can be called valid; and (4) an assessment's reliability is an estimate of the measurement task's accuracy or precision as a measurement instrument. The method of gathering evidence and characteristics of the candidate were identified as common sources of error associated with both objective and performance assessment. The following additional sources of error were associated with performance assessment: characteristics of the assessor; context of the assessment; and range and complexity of the task(s). It was concluded that each of these sources of error must be controlled throughout the assessment and that establishing clear task specifications (including evidence to be collected and decision-making rules) will increase reliability. (Contains 60 references.) (MN)

Clinical Assessment of Child and Adolescent Personality and Behavior

Understanding IEA's Comparative Studies of Student Achievement

An Investigation of the Reliability, Validity, and Usability of the Georgia Assessment of Music Achievement, Level B

Validity and Inter-Rater Reliability Testing of Quality Assessment Instruments

Reliability and Validity of International Large-Scale Assessment

The only text to comprehensively address the assessment of student learning in a wide variety of settings. Long considered the gold standard for evaluation and testing in nursing education, the sixth edition of this classic text provides expert, comprehensive guidance in the assessment of student learning in a wide variety of settings, as well as the evaluation of instructor and program effectiveness. It presents fundamental measurement and evaluation concepts that will aid nurse educators in the design, critique, and use of appropriate tests and evaluation tools. Important social, ethical, and legal issues associated with testing and evaluation also are explored, including the prevention of cheating and academic policies for testing, grading, and progression. Written by experts in the field of nursing education, Evaluation and Testing in Nursing Education features practical advice on the development of test blueprints; creation of all types of test items, including next-generation NCLEX-style items; the assembly, administration, and scoring of tests; test/item analyses and interpretation; evaluation of higher levels of learning; assessment of written assignments; and suggestions for creating tests in online courses and programs. An entire section is devoted to clinical evaluation processes and methods, including the use of simulation for assessment and high-stakes evaluation, clinical evaluation in distance settings, and the use of technology for remote evaluation of clinical performance. The text meets the National League for Nursing Certified Nurse Educator Competency #3: Use Assessment and Evaluation Strategies. NEW TO THE SIXTH EDITION: Expanded coverage of test item analysis and interpretation Expanded coverage of clinical evaluation processes and methods Guidance on how to work with part-time clinical educators and preceptors to ensure that evaluation processes are followed consistently Expanded content on the construction and administration of online tests Tips for adapting test bank items and other item-writing resources Guidelines for the design of academic policies for academic integrity, testing, grading, and progression

KEY FEATURES: Describes how to develop test blueprints and assemble, administer, write, and analyze tests Provides guidelines for the selection of standardized tests for a nursing curriculum Details how to evaluate written assignments with sample scoring rubrics Includes a robust ancillary package: Instructor's Manual (with sample syllabus, course modules, learning activities,

discussion questions, assessment strategies, and online resources) and chapter PowerPoint slides Purchase includes digital access for use on most mobile devices or computers

It is an essential skill for any strength and conditioning coach to be able to reliably assess the physical performance of their athletes and communicate the results and their implications to performers and coaches, alike. **Performance Assessment in Strength and Conditioning** is the first textbook to clearly and coherently suggest the most appropriate and reliable methods for assessing and monitoring athletes' performance, as well as including detailed sections on testing considerations and the interpretation and application of results. The book explores the full range of considerations required to reliably assess performance, including questions of ethics and safety, reliability and validity, and standardised testing, before going on to recommend (through a comparison of field- and laboratory-based techniques) the optimal methods for testing all aspects of physical performance, including: injury risk jump performance sprint performance change of direction and agility strength power aerobic performance body composition Closing with a section on interpreting, presenting and applying results to practice, and illustrated with real-life case study data throughout, **Performance Assessment in Strength and Conditioning** offers the most useful guide to monitoring athlete performance available. It is an essential text for upper-level strength and conditioning students and practitioners alike.

In the Fourth Edition of **Scale Development**, Robert F. DeVellis demystifies measurement by emphasizing a logical rather than strictly mathematical understanding of concepts. The text supports readers in comprehending newer approaches to measurement, comparing them to classical approaches, and grasping more clearly the relative merits of each. This edition addresses new topics pertinent to modern measurement approaches and includes additional exercises and topics for class discussion. Available with Perusall—an eBook that makes it easier to prepare for class Perusall is an award-winning eBook platform featuring social annotation tools that allow students and instructors to collaboratively mark up and discuss their SAGE textbook. Backed by research and supported by technological innovations developed at Harvard University, this process of learning through collaborative annotation keeps your students engaged and makes teaching easier and more effective. Learn more.

Theory and Applications

Environmental Valuation with Discrete Choice Experiments

Assessing the Reliability and Validity of the Competent Group

Communicator Problem Solving Assessment Instrument

How to Measure Survey Reliability and Validity

Validity and Inter-rater Reliability Testing of Quality Assessment Instruments

This book is open access under a CC BY-NC 2.5 license. This book describes the extensive contributions made toward the advancement of human assessment by scientists from one of the world's leading research institutions, Educational Testing Service. The book's four major sections detail research and development in measurement and statistics, education policy analysis and evaluation, scientific psychology, and validity. Many of the developments presented have become de-facto standards in educational and psychological measurement, including in item response theory (IRT), linking and equating, differential item functioning (DIF), and educational surveys like the National Assessment of Educational Progress (NAEP), the Programme of international Student Assessment (PISA), the Progress of International Reading Literacy Study (PIRLS) and the Trends in Mathematics and Science Study (TIMSS). In addition to its comprehensive coverage of contributions to the theory and methodology of educational and psychological measurement and statistics, the book gives significant attention to ETS work in cognitive, personality, developmental, and social psychology, and to education policy analysis and program evaluation. The chapter authors are long-standing experts who provide broad coverage and thoughtful insights that build upon decades of experience in research and best practices for measurement, evaluation, scientific psychology, and education policy analysis. Opening with a chapter on the genesis of ETS and closing with a synthesis of the enormously diverse set of contributions made over its 70-year history, the book is a useful resource for all interested in the improvement of human assessment.

Criterion-Referenced Test Development is designed specifically for training professionals who need to better understand how to develop criterion-referenced tests (CRTs). This important resource offers step-by-step guidance for how to make and defend Level 2 testing decisions, how to write test questions and performance scales that match jobs, and how to show that those certified as 'masters' are truly masters. A comprehensive guide to the development and use of CRTs, the book provides information about a variety of topics, including different methods of test interpretations, test construction, item formats, test scoring, reliability and validation methods, test administration, a score reporting, as well as the legal and liability issues surrounding testing. New revisions include: Illustrative real-world examples. Issues of test security. Advice on the use of test creation software. Expanded sections on performance testing. Single administration techniques for calculating reliability. Updated legal and compliance guidelines. Order the third edition of this classic and comprehensive reference guide to the theory and practice of organizational tests today.

Psychologists offer an increasing variety of services to the public. Among these services, psychological assessment of personality and behavior continues to be a central activity. One main reason is that other mental health professionals often do not possess a high level of competence in this area. And when dealing with children and adolescents, psychological assessment seems to take on an even greater role. Therefore, it follows that comprehensive graduate-level instruction in assessment should be a high priority for educators of psychologists who will work with these youth. This textbook is organized into three sections, consistent with the authors' approach to teaching. Part I provides students with the psychological knowledge base necessary for

modern assessment practice, including historical perspectives, measurement science, child psychopathology, ethical, legal, and cultural issues, and the basics of beginning the assessment process. Part II gives students a broad review of the specific assessment methods used by psychologists, accompanied by specific advice regarding the usage and strengths and weaknesses of each method. In Part III, we help students perform some of the most sophisticated of assessment practices: integrating and communicating assessment results and infusing assessment practice with knowledge of child development and psychopathology to assess some of the most common types of behavioral and emotional disorders in youth. A text focusing on assessment practices must be updated every four to six years to keep pace with advances in test development. For example, several of the major tests reviewed in the text, such as the Behavioral Assessment System for Children and the Child Behavior Checklist, have undergone major revisions since the publication of the last edition making the current content outdated. Further, another major test, the Conners' Rating Scales, is undergoing substantial revisions that should be completed before publication of the next edition. Finally, the evidence for the validity of the tests and the recommendations for their appropriate use evolve as research accumulates and requires frequent updating to remain current. For example, there was a special issue of the Journal of Clinical Child and Adolescent Psychology published focusing on evidenced-based assessment of the major forms of childhood psychopathology that will need to be integrated into the chapters in Part 3. This latter point reflects an important trend in the field that should influence the marketing of the book. That is, there are several initiatives being started in all of the major areas of applied psychology (e.g., school, clinical, and counseling) to promote evidenced-based assessment practices. These initiatives have all emphasized the need to enhance the training of graduate students in this approach to assessment. This has been the orientation of this textbook from its first edition: that is, Clinical Assessment of Child and Adolescent Personality and Behavior has focused on using research to guide all recommendations for practice. The ability of the textbook to meet this training need should be an important focus of marketing the book to training programs across all areas of applied psychology.

Issues of Reliability and Validity. Review of Research

Technical and Legal Guidelines for Corporate Training

Evaluation and Testing in Nursing Education, Sixth Edition

Reliability and Validity in Neuropsychological Assessment

Optimising New Modes of Assessment: In Search of Qualities and Standards

This book is an outcome of the Visual Values Workshop in 1982. It presents the ongoing research on state-of-the-art techniques and applications to address the human perception of changes in visual aesthetic resources and to assign psychological, social, and economic measures of value to visitors.

The present report focuses on the reliability and validity of combining subjective reports of symptomatology with objective reports of vomiting by experimenters in an expanded scale for motion sickness severity assessment.

Performance Improvement for Healthcare: Leading Change with Lean, Six Sigma, and Constraints Management

Reliability, Validity and Utility of the Functional Assessment Intervention Program

***The ITC International Handbook of Testing and Assessment
Reliability, Validity, and Responsiveness Ofthe Fall Risk Assessment Tool for
Subjects with Stroke
Assessment in Elementary and Secondary Education***