Online Assessment, Measurement and Evaluation

EMERGING PRACTICES

Online Assessment, Measurement, and Evaluation:

Emerging Practices

David D. Williams
Brigham Young University, USA

Scott L. Howell
Brigham Young University, USA

Mary Hricko
Kent State University, USA

Information Science Publishing

Hershey • London • Melbourne • Singapore

Acquisitions Editor: Renée Davies
Development Editor: Kristin Roth
Senior Managing Editor: Amanda Appicello
Managing Editor: Jennifer Neidig
Copy Editor: Maria Boyer
Typesetter: Cindy Consonery
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Publishing (an imprint of Idea Group Inc.)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@idea-group.com
Web site: http://www.idea-group.com

and in the United Kingdom by
Information Science Publishing (an imprint of Idea Group Inc.)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 3313
Web site: http://www.eurospan.co.uk

Copyright © 2006 by Idea Group Inc. All rights reserved. No part of this book may be reproduced, stored or distributed in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

Product or company names used in this book are for identification purposes only. Inclusion of the names of the products or companies does not indicate a claim of ownership by IGI of the trademark or registered trademark.

Library of Congress Cataloging-in-Publication Data

Online assessment, measurement, and evaluation : emerging practices / David D. Williams, Mary Hricko, and Scott L. Howell, editors.
p. cm.
Summary: "This book provides a view of the possibilities and challenges facing online educators and evaluators in the 21st century"--Provided by publisher.
Includes bibliographical references and index.
ISBN 1-59140-747-8 (hc) -- ISBN 1-59140-748-6 (sc) -- ISBN 1-59140-749-4 (ebook)
1. Computer-assisted instruction--Evaluation. 2. Educational tests and measurements--Data processing. 3. Distance education--Evaluation. I. Williams, David D., 1952- II. Hricko, Mary. III. Howell, Scott L.
LB1028.3.O552 2006
371.26'0285--dc22
2005013814

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

All work contributed to this book is new, previously-unpublished material. The views expressed in this book are those of the authors, but not necessarily of the publisher.

Preface

Although most of the 19 chapters in this volume were written in terms of measurement or assessment problems and issues, with little or no reference to evaluation as the larger context in which their meaning might be interpreted, each chapter will be briefly introduced in terms of an evaluation framework, which is developed further in Chapter I. In particular, the chapters after an introductory first chapter are grouped around approaches the authors took to collect evaluation data (Web surveys, student feedback, tests, and combinations), while other evaluation framework elements, such as stakeholders, evaluand (thing or person being evaluated), and criteria that are most relevant, are highlighted in each description.

Introduction to Measurement and Assessment in Evaluation

In "Measurement and Assessment Supporting Evaluation in Online Settings," David D. Williams of Brigham Young University outlines a framework for thinking about measurement and assessment in a more comprehensive evaluation context to help readers consider the other chapters in terms of an evaluation framework in hopes that they will see the common and unique contributions each chapter offers to their own evaluation (and related measurement and assessment) work. Reviewing key evaluation theorists, he notes that to be successful, evaluations must consider context, stakeholders, evaluand definitions, issues and concerns, values and criteria, questions, data collection and analysis, and reporting results, conclusions, and recommendations. He places measurement and assessment in this larger context to suggest a useful way of thinking about the rest of the chapters about online measurement and assessment.

Web Surveys

In "The Context and Culture of the Web as a Research Environment," Paula B. Doherty examines problems and how to resolve them when using self-administered surveys on the Web. Though she does not explicitly address the kinds of evaluation these surveys might be best used for, she discusses applied social research as a form of evaluation and identifies criteria for making surveys work via the Web. She notes the openness of the Web culture, couched in a helpful historical review of the Web, with implications for conducting surveys on the Web. Stakeholders are those the survey might relate to, and the evaluand is anything the survey might focus on, although she also examines the concept of a Web survey as an evaluand.

In "Computer-Adaptive Online Exit Surveys: Conceptual and Methodological Issues," Najmuddin Shaik proposes the creation and administration of adaptive surveys of dropouts from various higher education programs via online computers. In response to several stakeholders, such as administrators of universities, faculty, students, and particularly dropouts, Shaik explores the contexts of adaptive surveys and theories of dropouts in preparation for a survey approach. The chapter focuses on outlining the ideal characteristics of an instrument to help institutions identify dropout concerns as part of a formative evaluation process.

In "Quality Standards for Web-Based Learning: The Demand-Driven Learning Model," Krista Breithaupt and Colla J. MacDonald identify e-learning programs as generic evaluands and the Demand-Driven Learning Model (DDLM) as a clarification of criteria against which such learning should be evaluated for adult learner stakeholders. They develop a survey to evaluate how well e-learning programs meet the DDLM model and associated standards, and discuss data from three pilot studies.

In "Online Course-Ratings and the Personnel Evaluation Standards," Susan J. Clark, Christian M. Reiner, and Trav D. Johnson examine the use of online university course and faculty evaluations by students via Web surveys. The evaluand is the online formative and summative evaluation process. They use published personnel evaluation standards to meta-evaluate the assessment process, compare it to the traditional paper-based process, and recommend ways to improve it to the stakeholders: university faculty and administrators.

In "Case Study: Developing a University-Wide Distance Education Evaluation Program at the University of Florida," Christopher D. Sessums, Tracy A. Irani, Ricky Telg, and T. Grady Roberts explore the use of Web-based surveys to gather student evaluations. The stakeholders are faculty and administrators seeking summative and formative student feedback to improve distance courses. The chapter reviews the development of an instrument to be used throughout the university, with an emphasis on establishing validity and reliability based on concerns raised by 400 students through a qualitative Web survey, concerns from other institutions, accrediting bodies, faculty, and administrators.

In "Online Program Assessment: A Case Study of the University of Illinois at Urbana-Champaign Experience," Faye L. Lesht, Rae-Anne Montague, Vaughn J. Page, Najmuddin Shaik, and Linda C. Smith report on a university-wide initiative to evaluate off-campus graduate programs using online surveys in comparison with postal surveys. Combin-

试读结束，需要全本PDF请购买 www.ertongbook.com

ing online surveys to students and faculty with in-person interviews of administrators, the focus of the chapter is on formative program evaluation, with summative evaluation resulting as well. They report that better and more immediate communication with stakeholders and more economical use of resources resulted. Off-campus graduate students, university administrators, and faculty were the stakeholders represented by an independent committee established to evaluate off-campus programs.

Student Feedback

In "Cybercoaching: An Emerging Model of Personalized Online Assessment," Ni Chang and Naomi Jeffery Petersen explore the use of online e-mail and word processing to share teacher critique and coaching with university students regarding their writing. They conclude that the focus on formative development process and feedback to the students through coaching as teaching adheres to Vygotsky's social constructivism and is more efficient than many other forms of providing feedback. The main evaluands are students' learning and the quality of the student-teacher instructional experiences, as well as the Cybercoaching model. The authors illustrate use of this teaching-evaluation model from their own instructional experiences.

In "Testing the Validity of the Post and Vote Model of Web-Based Peer Assessment," Bruce L. Mann describes an approach he has developed to encourage university students to summatively evaluate other students' work using a Web-based peer assessment process. He presents validation data using instructor assessments as the standard. Students and faculty are the stakeholders most interested in the value of this model and the resulting performance evaluations. The chapter reviews the theoretical bases for the model and shares results of the validation process.

In "Online Assessment in a Teacher Education Program," Charles W. Peters and Patricia Ann Kenney describe and illustrate an online discussion and feedback process they use to give students formative and summative evaluative feedback via e-mail, "Track Changes" in Microsoft Word, and e-portfolios. As instructors, the authors analyze online discussion among their students, respond to and make online comments, and evaluate the quality of students' portfolios while adhering to theories of social constructivism and collaborative learning. Stakeholders and evaluators are student teachers and teacher educators who employ self-evaluation to study their contribution to the main evaluand: student teaching performance. They conclude that instruction based on a constructivist view of learning is easier using technology.

Tests

In "Variations in Adaptive Testing and Their Online Leverage Points," Roy Levy, John T. Behrens, and Robert J. Mislevy propose a taxonomy that differentiates assessments along three dimensions: (1) observation status, (2) claim status, and (3) locus of con-

trol. In detailing the taxonomy, they point out ways in which online assessment enables or enhances these features and highlight the inferential roles that adaptivity can play in assessment. They offer a principled perspective for examining advantageous features of various adaptive testing models, such as reduced time and increased precision in adaptive observation assessments and diagnostic capability in examinee-controlled assessments. Their views apply to a wide variety of stakeholders, evaluands, and kinds of evaluation that use adaptive tests.

In "Using Messick's Framework to Validate Assessment Tasks in Online Environments: A Course in Writing Effectively for UNHCR," Valerie Ruhe and Bruno D. Zumbo argue that various forms of technology (e.g., computerized testing) make old views of validity and assessment outmoded. They claim that Messick's framework brings psychometrics up to speed with these technologies. They illustrate their point with a study that employed surveys and interviews along with online feedback to students and tutors to evaluate a distance writing course for adults and the assessment task itself. They conclude that validation involves values and consequences, not just psychometric qualities of assessments. The online assessment tasks overlap with the value of the course components, and for this reason, online assessment tasks need to be validated in the broader context of the course as a system.

In "Online Assessment and Instruction Using Learning Maps: A Glimpse into the Future," Jim Lee, Sylvia Tidwell-Scheuring, and Karen Barton speculate about using online assessment to accelerate learning via effective links to instruction through quick assessment feedback and instruction generated by computers in response to scores. They examine notions of item adaptive testing, learning maps for instruction prescription, and validation of tests to enhance formative and summative evaluation of student understanding and associated instruction designed to help them through technology. Student (K-12) learning is the main evaluand. Instruction to help them is secondary. Teachers, students, parents, and administrators who are trying to help students learn are the main stakeholders, through tailoring of instruction based on review of learning maps and adaptive test results.

In "Authentic Assessment Online: A Practical and Theoretical Challenge in Higher Education," Smita Mathur and Terry Murray explore the use of online authentic assessment using electronic portfolios and journals, assessment embedded in online discussions, and associated rubrics to encourage continuous improvement of learning in an online environment. They build their approach around assumptions of social constructivism, and examine challenges and strategies to involve students and faculty in higher education as stakeholders evaluating learning in online classrooms.

In "Performance Testing: Validity Issues and Design Considerations for Online Testing," James B. Olsen reviews theoretical literature and builds on the premise that validation of tests should enhance evaluation of learning progress systems and students' learning. He argues for validation as evaluation through both performance and knowledge tests. Criteria for judging the quality of tests are dictated by the validity theory being used, and Olsen reviews two theories by Mislevy and Bunderson. These ideas apply to any situation involving performance assessment at any level and are relevant to stakeholders using tests as part of their evaluation process. Tests are the primary evaluands, and when they are evaluated positively, they can be used to evaluate student learning.

In "Assessment and College Progress: Capacity Building Through Formative and Summative Program Evaluation," Jennifer K. Holtz and Barbara Radner examine performance-based assessment in online courses at all levels, but particularly higher education. They define assessment as determining the relative value of something, in general, or an estimate of how close a student is to the knowledge and ability required, in particular. This assessment is used in formative and summative evaluation of students and courses. They argue that valid assessment is even more critical and challenging in online courses, the biggest challenge being accountability (Is the person doing the work the person getting the grade?). For the authors, the main stakeholders are teachers, students, and administrators, but this process should also enhance accreditation of institutions. They emphasize the linking of instruction, assessment, and integrated evaluation and using the online environment as a reason to enhance assessment. The chapter provides and illustrates tools for assessment in online courses, based on principles that apply to any educational program.

In "Evaluating Content-Management Systems for Online Learning Programs," Deborah L. Schnipke, Kirk Becker, and Jim S. Masters organized this chapter around summative evaluation questions test users and creators and organizations that develop, evaluate, and distribute tests should ask about online test item repositories or content management systems. The evaluands of interest are these Internet-based content-management systems, which can be used collaboratively by many participants to create tests. They discuss characteristics a content management system should provide for building tests relevant to any age group. Stakeholders are test and content management system users.

Combinations

In "Learning by Doing: Four Years of Online Assessment in Engineering Education Research," John C. Wise, Sang Ha Lee, and Sarah E. Rzasa present and reflect upon lessons learned through four study examples they engaged in at Penn State using different online tools. Using online surveys and tests, along with supporting qualitative measures, the authors explored the measurement of intellectual development, the evaluation of a minor within their college, the measurement of effects of an instructional method, and the establishment of validity and reliability using online data collection. Thus, evaluands included measurement instruments, intellectual development, educational programs, and instructional methods of both formative and summative interest to the students and faculty in their engineering college.

In "The Role of Assessment and Evaluation in Context: Pedagogical Alignment, Constraints, and Affordances in Online Courses," Julia M. Matuga distinguishes evaluation as making value judgments based on the assessments gathered in terms of the values of instructors, students, administrators, and other stakeholders. She explores the evaluation of higher education distance courses through the use of formative and summative quantitative measures of all kinds, along with qualitative feedback. Three experiences or cases are presented as context for a discussion of the unique challenges of using assessments to conduct evaluation in online settings.

Conclusion

Taken together with the chapters in the previous two volumes of this series on online measurement, assessment, and evaluation, the 19 chapters that follow provide a tantalizing view of the possibilities and challenges facing online educators and evaluators in the 21st century. As technology evolves and online measurement and assessment follow, the invitation of this volume is to use tried and tested evaluation principles to:

- employ these tools in evaluation systems that support stakeholders;
- clarify stakeholder's values and definitions of the evaluands they want to examine;
- help them think about the questions they most deeply want answered;
- keep the contexts and backgrounds in mind;
- seek to adhere to evaluation standards of feasibility, propriety, utility, and accuracy; and
- help participants realize that technical issues and methods are only of worth when they are in the service of helping people make thoughtful evaluation choices.

David D. Williams
January 2005

Acknowledgments

As senior editor for this volume, I thank and acknowledge my coeditors, Scott Howell and Mary Hricko, who invited me to join them on this third volume. Working with the 42 authors and 52 reviewers of the chapters included in this volume has been an educational and gratifying experience.

All the authors join me in acknowledging each of the 52 reviewers whose expertise, insights, and suggestions have added so much to the quality and value of the chapters in this volume. I have listed alphabetically by last name all reviewers who generously contributed to this volume: Stephanie Allen, Linda Barclay, Jane Birch, Rob Boody, Bruce Brown, Jeremy Browne, Holli Burgon, Olin Campbell, Suzy Cox, Richard Culatta, Joseph Curtin, Randy Davies, Yohan Delton, Dan Eastmond, Nick Eastmond, Deb Gearhart, Andy Gibbons, Abigail Gonzalez, Charles Graham, Mark Hawkes, Richard Helps, Scott Howell, Trav Johnson, Mike Johnson, Kristoffer Kristensen, Barbara Lawrence, Joseph Martineau, Paul Merrill, Deanna Molinari, David Nielson, Danny Olsen, David Pater, Courtney Peck, Tim Pelton, Ana Preto-Bay, Brian Radford, Duane Roberts, Larry Seawright, Sandip Sinharay, Joseph South, Steve Spencer, Todd Stubbs, Richard Sudweeks, Spencer Weiler, Rick West, Wendi Wilcken, Peter Williams, David Williamson, Todd Wilson, Steve Wygant, Stephen Yanchar, and Beverly Zimmerman.

Special appreciation is also due to Holly Baker, Laurel Lee Bingham, Chellae Brooks, Brianne Gillespie, Elisabeth Guyon, and Logan Molyenux of Brigham Young University, Division of Continuing Education, for the numerous and invaluable editing contributions they and their staff have provided to the chapter authors and volume editors. Special thanks also go to Jan Travers, managing editor for the publisher, and her team at Idea Group for their professional support and personal encouragement. I also appreciate my colleagues and family for supporting me in this endeavor.

David D. Williams, PhD
Brigham Young University
Provo, Utah, USA

Section I

Introduction to Measurement and Assessment in Evaluation

试读结束，需要全本PDF请购买 www.ertongbook.com

Online Assessment, Measurement, and Evaluation:

Emerging Practices

Table of Contents

Preface vi

Section I. Introduction to Measurement and Assessment in Evaluation

Chapter I. Measurement and Assessment Supporting Evaluation in Online Settings 1
David D. Williams, Brigham Young University, USA

Section II. Web Surveys

Chapter II. The Context and Culture of the Web as a Research Environment 10
Paula B. Doherty, Peninsula College, USA

Chapter III. Computer-Adaptive Online Exit Surveys: Conceptual and Methodological Issues 28
Najmuddin Shaik, University of Illinois at Urbana-Champaign, USA

Chapter IV. Quality Standards for E-Learning: The Demand-Driven Learning Model 45
Krista Breithaupt, American Institute for CPAs, USA
Colla J. MacDonald, University of Ottawa, Canada

Chapter V. Online Course-Ratings and the Personnel Evaluation Standards 61
Susan J. Clark, Brigham Young University, USA
Christian M. Reiner, Purdue University, USA
Trav D. Johnson, Brigham Young University, USA

Chapter VI. Case Study: Developing a University-Wide Distance Education Evaluation Program at the University of Florida 76
Christopher D. Sessums, University of Florida, USA
Tracy A. Irani, University of Florida, USA
Ricky Telg, University of Florida, USA
T. Grady Roberts, Texas A&M University, USA

Chapter VII. Online Program Assessment: A Case Study of the University of Illinois at Urbana-Champaign Experience 92
Faye L. Lesht, University of Illinois at Urbana-Champaign, USA
Rae-Anne Montague, University of Illinois at Urbana-Champaign, USA
Vaughn J. Page, University of Illinois at Urbana-Champaign, USA
Najmuddin Shaik, University of Illinois at Urbana-Champaign, USA
Linda C. Smith, University of Illinois at Urbana-Champaign, USA

Section III. Student Feedback

Chapter VIII. Cybercoaching: An Emerging Model of Personalized Online Assessment 110
Ni Chang, Indiana University - South Bend, USA
Naomi Jeffery Petersen, Indiana University - South Bend, USA

Chapter IX. Testing the Validity of the Post and Vote Model of Web-Based Peer Assessment 131
Bruce L. Mann, Memorial University, Canada

Chapter X. Online Assessment in a Teacher Education Program 153
Charles W. Peters, University of Michigan, USA
Patricia Ann Kenney, University of Michigan, USA

Section IV. Tests

Chapter XI. Variations in Adaptive Testing and Their Online Leverage Points 180
Roy Levy, University of Maryland, USA
John T. Behrens, Cisco Systems, Inc., USA
Robert J. Mislevy, University of Maryland, USA

Chapter XII. Using Messick's Framework to Validate Assessment Tasks in Online Environments: A Course in Writing Effectively for UNHCR 203
Valerie Ruhe, University of Maine, USA
Bruno D. Zumbo, University of British Columbia, Canada

Chapter XIII. Online Assessment and Instruction Using Learning Maps: A Glimpse into the Future 227
Jim Lee, CTB/McGraw-Hill, USA
Sylvia Tidwell-Scheuring, CTB/McGraw-Hill, USA
Karen Barton, CTB/McGraw-Hill, USA

Chapter XIV. Authentic Assessment Online: A Practical and Theoretical Challenge in Higher Education 238
Smita Mathur, University of South Florida, USA
Terry Murray, State University of New York at New Paltz, USA

Chapter XV. Performance Testing: Validity Issues and Design Considerations for Online Testing 259
James B. Olsen, Alpine Media Corporation, USA

Chapter XVI. Assessment and College Progress: Capacity Building Through Formative and Summative Program Evaluation 275
Jennifer K. Holtz, DePaul University School for New Learning, USA
Barbara Radner, DePaul University School for New Learning, USA

Chapter XVII. Evaluating Content-Management Systems for Online Learning Programs 292
Deborah L. Schnipke, Virtual Psychometrics, LLC, USA
Kirk Becker, University of Illinois and Promissor, USA
James S. Masters, University of North Carolina at Greensboro and Promissor, USA

Section V. Combinations

Chapter XVIII. Learning by Doing: Four Years of Online Assessment in Engineering Education Research 306
John C. Wise, The Pennsylvania State University, USA
Sang Ha Lee, The Pennsylvania State University, USA
Sarah E. Rzasa Zappe, The Pennsylvania State University, USA

Chapter XIX. The Role of Assessment and Evaluation in Context: Pedagogical Alignment, Constraints, and Affordances in Online Courses 316
Julia M. Matuga, Bowling Green State University, USA

About the Authors 331

Index 340

Chapter I

Measurement and Assessment Supporting Evaluation in Online Settings

David D. Williams, Brigham Young University, USA

Abstract

This chapter introduces an evaluation framework for interpreting online measurement and assessment as components of evaluation as an introduction to the other chapters in this volume. The evaluation framework includes attention to the evaluation context, stakeholders, evaluand (thing or person being evaluated), issues and concerns, values and criteria, questions, data collection and analysis, and reporting results, conclusions, and recommendations. This framework incorporates online measurement and assessment issues as important elements in evaluations of programs, personnel, and students.

Introduction

The following 18 chapters of this volume explore emerging practices in the use of online assessment and measurement to conduct evaluations of student learning (including tests, surveys, portfolios, and other assessments), educational programs, and personnel. Standards for judging the quality of evaluations in terms of utility, feasibility, propriety, and accuracy have been developed over the last 30 years by the Joint

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

Committee on Standards for Educational Evaluation (1984, 1988, 2003). Those who created these standards and many other independent scholars (e.g., Carey, 2001; Kubiszyn & Borich, 2003; Popham, 2000; Tanner, 2001; Thorkildsen, 2005; Ward & Murray-Ward, 1999) often use the words *assessment, measurement,* and *evaluation* as synonyms.

But when they describe the use of these three concepts, most authors appear to assume measurement is a means to assessment, and assessment is one important component of a much more complicated process called evaluation. Therefore, in this volume we differentiate among these terms to help readers appreciate the importance of all three concepts and how they can be used to improve measurement, assessment, evaluation, and the online learning and testing the volumes in this series are exploring.

Evaluation involves describing what is and what should be, and comparing the two. To gather information about what is, as well as what should be, assessment is an essential tool. And most assessments involve some kind of measurement process built upon theories of measurement. Scriven (1991), a noted evaluation theorist, summarizes each of these terms:

> *Measurement [is] a determination of the magnitude of a quantity, typically on a criterion-referenced test scale or on a continuous numerical scale. Whatever is used to do the measurement is called the measurement instrument. It may be a questionnaire or a test or an eye or a piece of apparatus. In certain contexts, we treat the observer as the instrument needing calibration or validation. Measurement is a common and sometimes large component of standardized evaluations, but a very small part of its logic, that is, of the justification for the evaluative conclusions.* (p. 266)

> *Assessment [is] often used as a synonym for evaluation in which the judgment [usually associated with evaluation] is built into the context of the numerical results. Raw scores on a test of no known content or construct validity would not be assessment; it is only when the test is—for example—of basic mathematical competence that reporting the results constitutes assessment in the appropriate sense, and of course the judgment of validity is the key evaluative component in this. Another part of the assessment movement, strongly supported in schools as well as colleges, is the move away from paper-and-pencil testing toward something more judgmental and global.* (p. 60)

> *The key sense of the term 'evaluation' refers to the process of determining the merit, worth, or value of something, or the product of that process. Terms used to refer to this process or part of it include: appraise, analyze, assess, critique, examine, grade, inspect, judge, rate, rank, review, study, test, measure. The evaluation process normally involves some identification of relevant standards of merit, worth, or value; some investigation of the performance of evaluands on these standards; and some integration or synthesis of the results to achieve an overall evaluation. It contrasts with the measurement process, which also involves the comparison of observations*

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

> *against standards, in that i) measurement is characteristically not concerned with merit, only with 'purely descriptive' properties, and ii) those properties are characteristically unidimensional, which avoids the need for the integrating step. The integration process is sometimes judgmental, sometimes the result of complex calculation, very commonly a hybrid of the two. In short [evaluation] is the sine qua non of intelligent thought and action, and in particular of professional practice.* (pp. 139-140)

More recently, these terms were defined by three different authors in the *Encyclopedia of Evaluation* (Mathison, 2005) as follows:

> Measurement *may be defined as the set of rules for transforming behaviors into categories or numbers. Constructing an instrument to measure a social science variable involves several steps, including conceptualizing the behaviors that operationally define the variable, drafting items that indicate the behaviors, administering draft items to try out samples, refining the instrument based on item analysis, and performing reliability and validity studies. These studies are necessary to ensure that scores on the instrument are consistent and have evidence of adequately representing a construct. Two theoretical approaches dominate the field of measurement: classical test theory and item response theory.* (Petrosko, 2005, p. 47)

> *From the Greek, 'to sit with',* assessment *means an evaluative determination. Roughly synonymous with testing and evaluation in lay terms, assessment has become the term of choice in education for determining the quality of student work for purposes of identifying the student's level of achievement. A more important distinction is between the terms assessment and measurement because educational constructs such as achievement, like most social phenomena, cannot be directly measured but can be assessed.* (Mabry, 2005, p. 22)

> Evaluation *is an applied inquiry process for collecting and synthesizing evidence that culminates in conclusions about the state of affairs, value, merit, worth, significance, or quality of a program, product, person, policy, proposal, or plan. Conclusions made in evaluations encompass both an empirical aspect (that something is the case) and a normative aspect (judgment about the value of something). It is the value feature that distinguishes evaluation from other types of inquiry, such as basic science research, clinical epidemiology, investigative journalism, or public polling.* (Fournier, 2005, pp. 139-140)

Synthesizing these views, although the terms are often used interchangeably, evaluation most often utilizes measurement and assessment to generate value judgments in a wide variety of situations. Assessment usually refers to evaluation of student learning, but

Copyright © 2006, Idea Group Inc. Copying or distributing in print or electronic forms without written permission of Idea Group Inc. is prohibited.

试读结束，需要全本PDF请购买 www.ertongbook.com