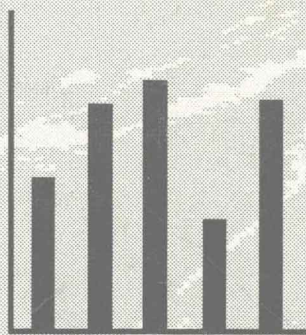


# **Developing Norm-Referenced Standardized Tests**



**Lucy Jane Miller**  
**Editor**



**The Haworth Press, Inc.**

# **Developing Norm-Referenced Standardized Tests**

Lucy Jane Miller  
Editor



The Haworth Press  
New York • London

*Developing Norm-Referenced Standardized Tests* has also been published as *Physical & Occupational Therapy in Pediatrics*, Volume 9, Number 1 1989.

© 1989 by The Haworth Press, Inc. All rights reserved. No part of this book may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, microfilm and recording, or by any information storage and retrieval system, without permission in writing from the publisher. Printed in the United States of America.

The Haworth Press, Inc., 10 Alice Street, Binghamton, NY 13904-1580  
EUROSPAN/Haworth, 3 Henrietta Street, London WC2E 8LU England

#### **Library of Congress Cataloging-in-Publication Data**

Developing norm-referenced standardized tests / Lucy Jane Miller, editor.

p. cm.

"Also . . . published as *Physical & occupational therapy in pediatrics*, volume 9, number 1, 1989"—T.p. verso.

Bibliography: p.

ISBN 0-86656-883-2

1. Occupational therapy—Research—Methodology. 2. Physical therapy—Research—Methodology. 3. Norm-referenced tests. I. Miller, Lucy J.

RM735.42.D38 1989

615.8'515—dc19

88-36502  
CIP

Dedicated to the memory of  
A. Jean Ayres

July 18, 1920 – December 16, 1988

Who took brilliant clinical insights and embodied them in a standardized format, defining them with self-imposed scientific discipline;

Who struggled to create and maintain a place within her chosen profession in which her work could flourish;

Who ultimately created a space for the rest of us where new and innovative ideas could grow within a supportive climate of excitement and scientific inquiry.

## **ABOUT THE EDITOR**

**Lucy Jane Miller, PhD, OTR**, Executive Director, The Foundation for Knowledge in Development, Englewood, Colorado, has been actively involved in early childhood research and test development for over 15 years. She is most widely recognized as the author of the Miller Assessment for Preschoolers (MAP), a nationally standardized, norm-referenced test. She is currently involved in two test development projects, one to develop a short developmental screening for preschoolers and the other to develop an assessment for infants and toddlers. Her research experience includes the standardization of the MAP and subsequent longitudinal predictive validity research, both large-scale national studies.

Dr. Miller holds a doctorate in early childhood special education, has published numerous technical and review papers, and was recently named a Charter Member to the American Occupational Therapy Foundation Academy of Research.



# Foreword

In the long run, men hit only what they aim at. Therefore, though they should fail immediately, they had better aim at something high.

— Thoreau  
“Economy” Walden (1854)

This collection of papers signals the early years of an important era, especially in occupational therapy. It reflects an attitude of professional and scientific responsibility that has been slow in coming but is now well on its way. A science is marked by the quality of and degree to which it measures the parameters of its field. Measuring instruments are critical tools for acquiring knowledge and it is difficult to acquire knowledge without them. The more precisely behavior is measured the better it is understood.

Tests yield numbers and numbers can do things that words or ideas cannot do. In occupational and physical therapy measurement is central to differential diagnosis, gain or loss assessment, establishing client status, predicting response to therapy, building and testing theory, and conveying information across fields. It is difficult to accomplish any of these goals without some form of measurement.

Recognition and acceptance of the need for measuring instruments is only the first step; the method of actual development of a test is described by this issue. The therapy fields face testing problems rather specific to themselves. Procedures that yield the most clinically meaningful information sometimes yield the least attractive statistical analysis. Therapists often measure functions that are

---

A. Jean Ayres, PhD, OTR, developed the *Sensory Integration and Praxis Test* and the *Southern California Sensory Integration Tests*. She is Emeritus member of the Board of Directors, Center for Study of Sensory Integrative Dysfunction, 5339 Bindewald Rd., Torrance, CA 90505.

not naturally distributed according to the Gaussian model on which most statistical methods in the behavioral sciences are based. Sometimes a linear transformation of raw scores into a different metric system is sufficient to meet the problem, but sometimes it is not. Occupational and physical therapists usually evaluate individuals whose performance falls below age expectations, yet most statistical procedures and expectations are based on a "normal" population.

Most importantly, the well-established behavioral sciences expect a quality of instrumentation the development of which requires large monetary expenditures, services of specialists in measurement, and cooperation on the part of many therapists. Meeting those expectations is often unrealistic, necessitating contentment with a more gradual course of development in which a test is designed, used, analyzed, and re-designed many times over many years in the clinical situation. It is a wise test-building method, for the design of a test is its most important attribute. When a test has proven its worth, it is then ready for a larger investment. The important thing is to measure the best way possible in a given situation. This volume will help accomplish that purpose.

*A. Jean Ayres*

## Preface

The great tragedy of science—the slaying of a beautiful hypothesis by an ugly fact.

—Thomas Henry Huxley (1825-1895)  
“*Biogenesis and Abiogenesis*”

*Developing Norm-Referenced Standardized Tests*, the title and theme of this special collection, has particular relevance for occupational and physical therapists. Its purpose is to provide a “how-to” manual for use by therapists who want to construct a *norm-referenced test*. In essence, this work will describe the process of developing standardized norm-referenced tests.

Members of both the occupational and physical therapy professions have increasingly engaged in the process of expanding and revising theories which undergird the practice of each profession, primarily by testing the hypotheses associated with specific theories. Clinicians in both professions also seek to develop tests that will enable them to (1) more precisely define the problems with which they are concerned and (2) select treatment procedures that will produce the most effective treatment outcomes.

The development of instrumentation enables the theoretician/scientist/clinician to describe and measure certain phenomena which are of interest and is a crucial component of knowledge development and scientific exploration and growth. In professions which are and have been practice-oriented since their founding, clinicians, as well as theoreticians and researchers, must develop an under-

---

Jerry A. Johnson, EdD, OTR, FAOTA, is President of Context, Inc., P.O. Box 6885, Denver, CO 80206. Formerly, she served as President of the American Occupational Therapy Association and as Professor of Occupational Therapy at Boston University and the Washington University in St. Louis.



standing of the importance of research and its potential contributions to the improvement of their practice.

Two types of tests, *norm-referenced tests* and *criterion-referenced tests*, are relevant for the clinician who is beginning research. A norm-referenced test is one in which an individual's performance is compared against the performance of other individuals (a referent or normative group). Most common among this type of tests are intelligence or aptitude tests such as the *Bayley Scales of Infant Development*,<sup>1</sup> *Wechsler Intelligence Scale for Children-Revised*,<sup>2</sup> and the *McCarthy Scales of Children's Abilities*.<sup>3</sup>

A criterion-referenced test compares an individual's performance to a predetermined behavioral criterion or content domain.<sup>4</sup> Criterion-referenced tests are used primarily in education, and include tests assessing basic mastery of skills, such as the *Metropolitan Achievement Test, Sixth Edition*,<sup>5</sup> and the *Criterion-Referenced Curriculum-Mathematics*.<sup>6</sup> In the therapies there are several criterion-referenced checklists such as the *Developmental Programming for Infants and Young Children*<sup>7</sup> and the *Vulpe Assessment Battery*.<sup>8</sup> A criterion-referenced test indicates whether or not the individual has learned certain skills and may describe which skills are learned.

In summary, a criterion-referenced test reports the examinee's performance in terms of what s/he knows and can do, whereas the norm-referenced test reports how the examinee's performance compares to the performance of others. One other difference in these two types of tests is that norm-referenced tests tend to be standardized, whereas criterion-referenced tests are not necessarily standardized.<sup>9</sup> While the content of this volume only discusses *norm-referenced tests*, both types of tests are important but in different ways. Table 1 describes the differences between norm-referenced tests and criterion-referenced tests.

This special collection is designed for therapists who have an interest in conducting research, either collaboratively with established scientists or independently, to pursue questions of interest. More specifically, the content of this issue will take the reader through a step-by-step process, including identification of a concept that should be subjected to testing, development of appropriate test items, and the procedures for standardizing a norm-referenced test.

In Chapter 1, the reader is introduced to the historical perspective

Table 1. Differences Between Norm-Referenced Tests and Criterion-Referenced Tests.

	Norm-Referenced Tests	Criterion-Referenced Tests
Purpose	Comparison among examinees	Examinee performance in relation to set of competencies
Content Specificity	<ul style="list-style-type: none"> <li>• More general content</li> <li>• Often greater breadth of coverage</li> </ul>	<ul style="list-style-type: none"> <li>• More detailed content specifications</li> <li>• May use behavioral or instructional objectives</li> </ul>
Test Development	<ul style="list-style-type: none"> <li>• Use of item statistics (difficulty and discrimination indexes) serve in item selection.</li> <li>• Select moderate difficulty items and high discrimination.</li> </ul>	Content includes items passed or failed by nearly all examinees.
Test Score Generalizability	<ul style="list-style-type: none"> <li>• Generally do not make generalizations from norm-referenced achievement test scores.</li> </ul>	Generalize performance to larger domains of content defining the competencies.

Table by Sharon Cernak, EdD, OTR

of and projected trends associated with test development. The rationale of standardized testing is discussed, as is the need for involvement of occupational and physical therapists in test development.

Chapter 2 assists the reader in the formulation of ideas leading to construction of an appropriate test. The components of planning the development process are elaborated, from the review of the theoretical and research literature to the resources available to fund stages of the research.

In Chapter 3, the importance and step-by-step details of the Development Edition phase are illuminated. During this stage, administration procedures are clarified and item analysis is conducted to determine whether the items have good psychometric characteristics. A large item pool is generated and studied extensively to determine which items should be selected for large scale standardization.

Procedures for standardization of the test, including development of a scoring system and testing for reliability and validity, are discussed in Chapter 4, 5, 6, and 7. Chapter 8 outlines methods of preparing an examiner's manual and presentation of its content. A "personal note" from the editor of this special collection is also included in which she recounts experiences as a test developer.

Standardization of tests in a profession forms the foundation for its credibility and begins to delineate the parameters of practice and expertise which clinicians, educators, and/or researchers in other disciplines can expect of its members. In effect, practice based on results and information obtained from standardized tests developed for specific purposes establishes a profession's claim to certain areas of knowledge, practice, and responsibility. It moves the profession's practice from a "belief" system to a foundation which others can evaluate objectively.

The authors of this issue are making a significant contribution to the continuing development of knowledge and research in both occupational and physical therapy. Not only will this increase understanding of the process of test development for instruments which are already available, but it is intended to stimulate the development of new instrumentation by providing a step-by-step guide for those interested in developing a new test. Developing more standardized tests in the therapies will contribute to the continuing intellectual growth of members of these two professions, and to greater public

recognition and acceptance of their theoretical foundations and therapeutic procedures.

This work is timely and is recommended for study by all occupational and physical therapists who want to actively participate in the challenges that confront their respective professions and who desire to pursue their own professional growth.

Jerry A. Johnson

## REFERENCES

1. Bayley N: *Bayley Scales of Infant Development*. New York, Psychological Corporation, 1969.
2. Wechsler D: *Wechsler Intelligence Scale for Children-Revised*. New York, Psychological Corporation, 1974.
3. McCarthy D: *McCarthy Scales of Children's Abilities*. New York, Psychological Corporation, 1972.
4. Anastasi A: *Psychological Testing*, ed 6. New York, Collier Macmillan Publishers, 1988.
5. Farr R, Prescott G, Balow I, Hogan T: *Metropolitan Achievement Tests, Sixth Edition-Reading Diagnostic Tests*. New York, Psychological Corporation, 1987.
6. Stephens T: *Criterion-Referenced Curriculum-Mathematics*. New York, Psychological Corporation, 1982.
7. Rogers SJ, D'Eugenio DB, Moersch M: *Developmental Programming for Infants and Young Children*. Ann Arbor, MI, University of Michigan Press, 1977.
8. Vulpe SG: *Vulpe Assessment Battery*, ed 2. Toronto, National Institute on Mental Retardation, 1977.
9. Goodwin WL, Driscoll L: *Handbook for Measurement and Evaluation in Early Childhood Education*. San Francisco, Jossey-Bass Publishers, 1980.

# Developing Norm-Referenced Standardized Tests

## CONTENTS

<b>Foreword</b>	<b>xi</b>
<i>A. Jean Ayres</i>	
<b>Preface</b>	<b>xiii</b>
<i>Jerry A. Johnson</i>	
<b>Acknowledgements</b>	<b>xix</b>
<b>Chapter 1: Measurement in Developmental Therapy: Past, Present, and Future</b>	<b>1</b>
<i>Suzann K. Campbell</i>	
Introduction	1
History of Standardized Testing	2
Lack of Testing Instruments	5
Rationale for Increased Use of Standardized Tests	8
The Priority to Develop Validated Tools	9
Summary and Recommendations	10
<b>Chapter 2: Planning the Initial Version</b>	<b>15</b>
<i>Bette R. Bonder</i>	
Introduction	15
Formulating the Idea	16
Reviewing the Literature	19
Identification of Specific Content for Items	20
The Table of Specifications	21
Item Format	22
Item Development	28
Planning the Standardization Process	32
Statistical Consultation	32

Ethical Considerations	33
Funding the Plan	37
Summary and Recommendations	39
<b>Chapter 3: Designing and Implementing Research on the Development Editions of the Test</b>	<b>43</b>
<i>Jan Gwyer</i>	
Introduction	43
Purpose for Conducting a Development Edition	45
Selecting the Development Edition Sample	46
Test Administration Procedures	47
Analysis of Item Characteristics	52
Evaluating Reliability and Validity	56
Summary and Recommendations	59
<b>Chapter 4: Standardizing an Assessment</b>	<b>63</b>
<i>James Gyurke</i> <i>Aurelio Prifitera</i>	
Introduction	63
Developing the Timeline and Budget	64
Examiner Recruitment and Training	69
Selecting a Sample	70
Bias Review	77
Data Collection	78
Data Analysis	80
Reliability and Validity	85
Norms Development	86
Publication	86
Summary and Recommendations	87
<b>Chapter 5: Norms and Scores</b>	<b>91</b>
<i>Sharon Cermak</i>	
Introduction	91
Statistical Concepts	92
Norms and Scores	98
Summary and Recommendations	113



<b>Chapter 6: Reliability</b>	<b>125</b>
<i>Jean C. Deitz</i>	
Introduction	125
Types of Reliability	126
Designing Reliability Research for a Test	135
Statistics	137
Summary and Recommendations	144
<b>Chapter 7: Validity</b>	<b>149</b>
<i>Winnie W. Dunn</i>	
Introduction	149
Validity in Therapeutic Practice	150
Basic Issues in Validity Research	151
Types of Validity Studies	153
Summary and Recommendations	164
<b>Chapter 8: Preparing the Examiner's Manual</b>	<b>169</b>
<i>Tracy A. Sprong</i>	
Introduction	169
Preliminary Preparation	170
Getting Started	171
The Drafting Process	172
Final Organization	175
Summary and Recommendations	179
<b>Epilogue: Test Development on the Installment Plan     or "How I Developed a Test in 27,000 Easy Steps"</b>	<b>185</b>
<i>Lucy Jane Miller</i>	
<b>Appendixes</b>	<b>195</b>

those tests that are available to work in improving clinical diagnostic capabilities and accountability to clients and the public for the costly treatment provided.

The purpose of this introductory chapter is to describe the history of standardized testing, the current problem of lack of testing instruments, and the rationale for increased use of standardized tests in developmental therapy settings. The need for pediatric occupational and physical therapists to develop more scientifically validated tools is also addressed.

### ***HISTORY OF STANDARDIZED TESTING***

Critical to the future development of occupational and physical therapy as scientific and clinical disciplines is the generation of measurement tools specific to client needs. Many clinicians are depending upon tests developed by psychologists and other professionals with different types of clients or goals in mind. Worse, a trend of the 1970s in some areas of the subdisciplines was to avoid formal testing.

According to Nunnally, the movement toward standardized testing was generated in the last century by Darwin's theory of evolution.<sup>4</sup> The concept of "survival of the fittest" led to interest in measuring individual differences in abilities of various types. Galton, the founder of the eugenics movement aimed at improving the human race, studied the heritability of human traits using tests of sensory discrimination. Sensory acuity was believed to be the basis of intellect. Interest was so great that people were willing to pay for the opportunity to be measured with Galton's techniques! His contributions were important for their explicit recognition of (1) the need for standardization in testing—the concept that each individual should be tested with the same items under the same conditions and with standard instructions, (2) for emphasizing the importance of individual differences in abilities, and (3) for the development of correlational methods, later refined by Pearson, with which to analyze the collected data.

Binet and Simon developed the first test of global intelligence in 1905 at the behest of the French government which had recognized the need for developing a testing tool to evaluate and classify chil-

dren who were too mentally deficient to profit from schooling.<sup>4</sup> This work led to the concept of norms for performance of children at different ages, calculation of mental ages, and eventually the search for factors in human intelligence. Psychologists, such as Spearman and Thurstone, theorized that intelligence included a general factor, *g*, and specific factors, such as verbal, numerical, spatial, memory, reasoning, analogy, and perceptual abilities. They developed factor analysis as a methodological approach to studying human cognitive abilities. Piaget further advanced the study of cognitive performance by carefully studying the processes, rather than the content, of mental development. This produced a revolutionary and continuing impact on developmental psychology.

The standardized testing movement spread widely to include assessment of most areas of human ability, and made significant contributions in practical application to personnel selection, school admissions, and psychiatric and other medical diagnostic tasks.<sup>4</sup> Rare is the person who has never taken a standardized test of some kind before reaching adolescence. Although standardized tests are continually criticized for labeling and lack of cultural validity for some groups in the population, such as minorities, they remain the best known means of sorting, classifying, diagnosing, and measuring progress.

Researchers have also studied the factors contributing to motor development, to motor learning, and to skilled motor performance. Developmental therapists have been interested in studying the relationships between motor milestones, developmental reflexes and reactions, and “quality” of movement. Though interested, little advancement has been made in the measurement of motor dysfunction or progress during therapy. Undoubtedly, progress has been slow in part because of the tremendous complexity involved in sorting out the many factors, both sensory and motor, that contribute to skilled motor performance. “Quality” of movement is difficult to capture and describe because it does not consist of a single factor, but rather is a jargon term inclusive of coordination, postural control, and balance.<sup>5</sup>

Research on motor development has resulted in the identification of nine different aspects of gross and fine motor development in normal children from preschool age to adolescence.<sup>6</sup> Specific fac-