

DOCUMENT RESUME

ED 378 394

CE 068 055

AUTHOR Murray, T. Scott
TITLE Proxy Measurement of Adult Basic Skills: Lessons from Canada.
INSTITUTION National Center on Adult Literacy, Philadelphia, PA.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO NCAL-TR94-18
PUB DATE Jan 95
CONTRACT R117Q00003
NOTE 32p.
AVAILABLE FROM National Center on Adult Literacy, Dissemination, 3910 Chestnut Street, Philadelphia, PA 19104-3111 (order no. TR94-18: \$7).
PUB TYPE Viewpoints (Opinion/Position Papers, Essays, etc.) (120) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Adult Basic Education; *Adult Literacy; *Basic Skills; Developed Nations; Estimation (Mathematics); Evaluation Methods; *Evaluation Utilization; Foreign Countries; Literacy Education; Measurement Techniques
IDENTIFIERS *Canada; *Proxy Measurement

ABSTRACT

"Proxy" measures have been developed and applied that extend the usefulness of direct assessments of adult basic skills. Three categories of users of data on adult literacy are as follows: (1) government authorities who need objective data to inform the decision-making process; (2) program delivery institutions who require an assessment tool to diagnose individual student needs, information to guide their own program and product development, and information to justify government funding; and (3) individuals who require information to define their own need for remedial training and to marshal societal awareness and maintain the political will to justify continuing investment. Prevailing approaches to measurement have been criticized for several reasons: a common definition of literacy is not feasible or desirable and the estimates derived from statistics are useless due to measurement error. The Canadian experience offers insight into the implications of direct assessments for proxy measures in three areas: correction factors for estimates of educational attainment-based proxy measures, identification of new proxy measures that could be collected, and proxy skills for small areas. The availability of new and improved proxy measures can then affect the design and output of the direct assessments themselves. Direct assessments could be administered at greater intervals; the basic objectives should be reconsidered; and more informed multivariate analysis is needed. Appendixes include three data tables. Contains 27 references. (YLB)



NATIONAL CENTER ON ADULT LITERACY

**PROXY MEASUREMENT OF ADULT
BASIC SKILLS:**

LESSONS FROM CANADA

T. Scott Murray
Statistics Canada

NCAL TECHNICAL REPORT TR94-18
JANUARY 1995

U.S. DEPARTMENT OF EDUCATION
Office of Educational Resources and Information
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

This document has been reproduced as
received from the person or organization
originating it

Minor changes have been made to
improve reproduction quality

• Points of view or opinions stated in this
document do not necessarily represent
official OERI position or policy

02058 055

**PROXY MEASUREMENT OF ADULT
BASIC SKILLS:**

LESSONS FROM CANADA

T. Scott Murray
Statistics Canada

NCAL TECHNICAL REPORT TR94-18
JANUARY 1995

This work was presented at a roundtable cosponsored by the Organisation for Economic Co-operation and Development (OECD) and the National Center on Adult Literacy (NCAL) at the University of Pennsylvania. NCAL is part of the Educational Research and Development Center Program (Grant No. R117Q00003) as administered by the Office of Educational Research and Improvement, U.S. Department of Education, in cooperation with the Departments of Health and Human Services and Labor. The findings and opinions expressed here do not necessarily reflect the position or policies of the OECD, NCAL, or the U.S. Department of Education.

NATIONAL CENTER ON ADULT LITERACY
UNIVERSITY OF PENNSYLVANIA
3910 CHESTNUT STREET
PHILADELPHIA, PA 19104-3111
PHONE (215) 898-2100 FAX (215) 898-9804

The National Center on Adult Literacy (NCAL) was established in 1990 by the U.S. Department of Education, with co-funding from the Departments of Labor and Health and Human Services. The mission of NCAL addresses three primary challenges: (a) to enhance the knowledge base about adult literacy; (b) to improve the quality of research and development in the field; and (c) to ensure a strong, two-way relationship between research and practice. Through applied research and development and dissemination of the results to researchers, policymakers, and practitioners, NCAL seeks to improve the quality of adult literacy programs and services on a nationwide basis. NCAL serves as a major operating unit of the Literacy Research Center at the University of Pennsylvania.

NCAL publications to date include:

- May 1992 *Matching Literacy Testing With Social Policy: What are the Alternatives?*
Richard L. Venezky (PB92-01, 7 pages)
- Oct 1992 *Life-Span and Life-Space Literacy: Research and Policy in National and International Perspectives*
Daniel A. Wagner (OP92-01, 15 pages)
- Oct 1992 *Expanding Theories of Adult Literacy Participation*
Karen Reed Wikelund, Stephen Reder, & Sylvia Hart-Landsberg (TR92-01, 30 pages)
- Oct 1992 *Invitations to Inquiry: Rethinking Staff Development in Adult Literacy Education*
Susan L. Lytle, Alisa Belzer, & Rebecca Reumann (TR92-02, 44 pages)
- Dec 1992 *Developing the Professional Workforce for Adult Literacy Education*
Susan L. Lytle, Alisa Belzer, & Rebecca Reumann (PB92-02, 11 pages)
- Jan 1993 *The Analysis of Adult Literacy: Problems in Factor Analysis Bib-Spiralled Item Administration*
David Kaplan (OP93-01, 18 pages)
- Mar 1993 *The Impact of Workplace Literacy Programs: A New Model for Evaluation of Workplace Literacy Programs*
Larry Mikulecky & Paul Lloyd (TR93-02, 180 pages)
- Mar 1993 *Literacy and Machines: An Overview of the Use of Technology in Adult Literacy Programs*
Terilyn C. Turner (TR93-03, 86 pages)
- Jun 1993 *Myths and Misconceptions in Adult Literacy: A Research and Development Perspective*
Daniel A. Wagner (PB93-01, 10 pages)
- Jun 1993 *Literacy and Development: Rationales, Assessment, and Innovation*
Daniel A. Wagner (IP93-01, 50 pages)
- Jun 1993 *Early Childhood, Family, and Health Issues in Literacy: International Perspectives*
Laurel D. Puchner (IP93-02, 45 pages)
- Sep 1993 *What Makes Worker Learn? The Role of Incentives in Workplace Education and Training*
Donald Hirsch & Daniel A. Wagner (Eds.) (IP93-03, 243 pages)
- Sep 1993 *Prison Literacy: Implications for Program and Assessment Policy*
Anabel Newman, Warren Lewis, & Carolyn Beverstock (TR93-01, 219 pages)
- Sep 1993 *Management Information Systems in Adult Education: Perspectives from the States and from Local Programs*
Mark A. Kutner, Lenore Webb, Rebecca Herman, & Pelavin Associates, Inc. (TR93-04, 150 pages)
- Sep 1993 *What Can Employers Assume about the Literacy Skills of GED Graduates?*
David Kaplan & Richard L. Venezky (TR93-05, 45 pages)

NCAL publications to date (continued)

- Sep 1993 *Should Reading-Disabled Adults Be Distinguished From Other Adults Seeking Literacy Instruction? A Review of Theory and Research*
Anne E. Fowler & Hollis S. Scarborough (TR93-07, 101 pages)
- Sep 1993 *When Less Is More: A Comparative Analysis for Placing Students in Adult Literacy Classes*
Richard L. Venezky, Page S. Bristow, & John P. Sabatini (TR93-08, 46 pages)
- Sep 1993 *Metacognitive Aspects of Adult Literacy*
Scott G. Paris & Andrea Parecki (TR93-09, 44 pages)
- Nov 1993 *Teamwork and Literacy: Learning from a Skills-Poor Position*
Sylvia Hart-Landsberg & Steve Reder (TR93-06, 63 pages)
- Nov 1993 *Motivations for Learning: Voices of Women Welfare Reform Participant*
Karen Wikelund (TR93-10, 54 pages)
- Nov 1993 *Initiating Practitioner Inquiry: Adult Literacy Teachers, Tutors, and Administrators Research Their Practice*
Susan L. Lytle, Alisa Belzer, & Rebecca Reumann (TR93-11, 69 pages)
- Nov 1993 *Coalition Building for Adult Literacy: Historical and Organizational Perspectives*
Anabel P. Newman & Bernadette Lehman (TR93-13, 68 pages)
- Nov 1993 *Effective Service Delivery in Adult Literacy Programs: A Policy Review and Recommendations*
Judith Ann Koloski (TR93-14, 46 pages)
- Dec 1993 *Issues and Challenges in Adult Numeracy*
Iddo Gal (TR93-15, 62 pages)
- Dec 1993 *Adult Literacy Training and the Integration of Human Services*
Elizabeth R. Reisner (TR93-16, 30 pages)
- Apr 1994 *Measuring Gain in Adult Literacy Programs*
Richard L. Venezky, Page S. Bristow, & John P. Sabatini (TR93-12, 24 pages)
- Apr 1994 *Understanding Family Literacy: Conceptual Issues Facing the Field*
Vivian L. Gadsden (TR94-02, 32 pages)
- Apr 1994 *Children, Parents, and Families: An Annotated Bibliography on Literacy Development In and Out of Program Settings*
Vivian L. Gadsden, Ludo C. P. Scheffer, & Joel Hardman (TR94-04, 84 pages)
- Jun 1994 *Literacy Transfer: A Review of the Literature*
Larry Mikulecky, Peggy Albers, & Michele Peers (TR94-05, 21 pages)
- Jun 1994 *Instruction and Assessment for Limited-English-Proficient Adult Learners*
Ronald W. Solórzano (TR94-06, 33 pages)
- Jun 1994 *Early Warning Signs of Functional Illiteracy: Predictors in Childhood and Adolescence*
Nazli Baydar, Jeanne Brooks-Gunn, & Frank F. Furstenberg (OP94-01, 13 pages)
- Jul 1994 *Use It or Lose It? The Problem of Adult Literacy Skill Retention*
Daniel A. Wagner (TR94-07, 27 pages)
- Jul 1994 *Technology: New Tools for Adult Literacy, Videoconference Participant Materials*
Joyce Harvey-Morgan, Chris Hopey, & Karl Rethemeyer (Eds.) (PR94-01, 58 pages)
- Sep 1994 *Supply and Demand for Literacy Instruction in the United States*
Richard L. Venezky & Daniel A. Wagner (TR94-10, 13 pages)
- Sep 1994 *The Professionalization of the Teacher in Adult Literacy Education*
Timothy Shanahan, Maureen Meehan, & Stephen Mogge (TR94-11, 20 pages)
- Sep 1994 *The Role of Literacy in the Wealth of Individuals and Nations*
Suc E. Berryman (TR94-13, 15 pages)

NCAL publications to date (continued)

- Oct 1994 *Abilities and Competencies in Adulthood: Life-Span Perspectives on Workplace Skills*
Jacqui Smith & Michael Marsiske (TR94-12, 36 pages)
- Nov 1994 *Literacy and Adults With Developmental Disabilities*
Karen A. Erickson, David A. Koppenhaver, & David E. Yoder (TR94-15, 31 pages)
- Nov 1994 *Adult Basic Skills in OECD Countries: Policy Issues and a Research Agenda*
David Stern & Albert Tuijnman (IP94-01, 12 pages)
- Dec 1994 *Who Counts in Adult Literacy Programs? A National Survey of Numeracy Education*
Iddo Gal & Alex Schuh (TR94-09, 20 pages)
- Dec 1994 *Literacy and Welfare Reform: Are We Making the Connection?*
Elena Cohen, Susan Golonka, Rebecca Maynard, Theodora Ooms, & Todd Owen
(TR94-16, 47 pages)
- Jan 1995 *Literacy and Older Adults in the United States*
Gail Weinstein-Shr (TR94-17, 25 pages)
- Jan 1995 *Self-Assessed Skill Needs and Job Performance*
Peter Capelli & Nikolai Rogovsky (TR94-08, 12 pages)
- Jan 1995 *Proxy Measurement of Adult Basic Skills: Lessons From Canada*
T. Scott Murray (TR94-18, 18 pages)
- Jan 1995 *Using Large-Scale Assessment Results to Identify and Evaluate Generalizable Indicators of Literacy*
Irwin S. Kirsch & Ann Jungeblut (TR94-19, 14 pages)

Information on ordering of NCAL publications may be addressed to Dissemination at NCAL.
Revised January 16, 1995

NCAL MANAGEMENT

Daniel A. Wagner, Director
Richard L. Venezky, Co-Director for Research and Development
Joyce Harvey-Morgan, Associate Director
Vivian L. Gadsden, Associate Director
Sandra K. Stewart, Manager of Dissemination
Mary O. Russell, Administrative Coordinator
Janet C. Smith, Editor
R. Karl Rethemeyer, Manager, Literacy Technology Laboratory

NCAL SENIOR PROJECT DIRECTORS

María Carlo, University of Pennsylvania
Vivian L. Gadsden, University of Pennsylvania
Ido Gal, University of Pennsylvania
Joyce Harvey-Morgan, University of Pennsylvania
Susan L. Lytle, University of Pennsylvania
Larry Mikulecky, Indiana University
Scott G. Paris, University of Michigan
Laurel D. Puchner, University of Pennsylvania
Stephen Reder, Northwest Regional Educational Laboratory
R. Karl Rethemeyer, University of Pennsylvania
Regie Stites, University of Pennsylvania
Richard L. Venezky, University of Delaware
Daniel A. Wagner, University of Pennsylvania

NCAL NATIONAL ADVISORY PANEL

Chair: Gloria Twine Chisum, Vice-Chair. University of Pennsylvania Board of Trustees
Richard C. Anderson, Director. Center for the Study of Reading. University of Illinois
Joan D. Baraloto, Director. Education and Family Initiatives. USA Today
Jack Bowsher, Director of Education (ret.). IBM. Inc.
Jeanne Chall, Professor. Graduate School of Education. Harvard University
Lillian Escobar-Haskins, Executive Director. Governor's Advisory Commission on Latino Affairs. Pennsylvania
The Honorable William F. Goodling, U.S. Representative. Pennsylvania
Carl Kaysen, David W. Skinner Chair of Political Economy. Massachusetts Institute of Technology
Gerald Kilbert, Director of Adult Education. California
Luis Laosa, Principal Research Scientist. Research Department. Educational Testing Service
Noreen Lopez, Manager. Adult Education Product Development. Contemporary Books. Inc.
Geraldine Novelo, Deputy Director (ret.). Instituto Nacional para la Educación de los Adultos. Mexico
Van D. Ooms, Senior Vice President and Director of Research. Committee for Economic Development
Bernard P. Reca, Vice President. Bell Atlantic
Anthony Sarmiento, Assistant Director. Human Resources Development Institute. AFL-CIO
Robert Schwartz, Program Director. Education. Pew Charitable Trusts
Ramsay Selden, Director. State Education Assessment Center. Council of Chief State School Officers
Senator Paul Simon, U.S. Senator. Illinois
Dorothy Strickland, Professor. Graduate School of Education. Rutgers University
Francis X. Sutton, Vice President (ret.). Ford Foundation
Calvin Tyler, Senior Vice President for Human Resources. UPS. Inc.
Peter Waite, Executive Director. Laubach Literacy Action

NCAL EDITORIAL REVIEW COMMITTEE

Hal Beder, Rutgers University
María Carlo, University of Pennsylvania
Suzanne Cockley, Virginia Adult Education Research Network
Aydin Durgunoglu, University of Illinois. Urbana
Marcia Farr, University of Illinois at Chicago
Beth Foley, Utah State University
Maggie Gaines, Baltimore City Literacy Corp.
Sheryl Gowen, Georgia State University
Keiko Koda, Ohio University
Susan L. Lytle, University of Pennsylvania
Peter B. Mosenthal, Syracuse University
Thomas Reeves, University of Georgia
Timothy Shanahan, University of Illinois at Chicago
Keith Stanovich, University of Toronto
Sondra Stein, National Institute for Literacy
Terrence G. Wiley, California State University Long Beach

ACKNOWLEDGMENTS

I would like to acknowledge the seminal work of David Neice and Margaret Adsett, which this report extends. Thanks are due to Stan Jones and Irwin Kirsch, who taught me the subtleties of measuring adult literacy, and to Richard Shillington for his multivariate prowess.

TABLE OF CONTENTS

<i>Acknowledgments</i>	<i>i</i>
<i>Table of Contents</i>	<i>iii</i>
<i>Abstract</i>	<i>v</i>
Introduction	1
Proxy Measures Defined	1
A Brief History of Proxy Measurement in Adult Literacy	2
The Uses of Data on Adult Literacy	2
Adult Literacy Defined	4
Criteria for Judging the Utility of Proxy Measures	5
Problems With Prevailing Approaches to Measurement	6
The Advent of Direct Skill Assessments in North America	8
Implications of Direct Assessments for Proxy Measures of Adult Literacy	11
Implications of Proxy Measures for Direct Assessments	14
Conclusion	16
<i>References</i>	<i>17</i>
<i>Appendix: Tables</i>	<i>A-i</i>

PROXY MEASUREMENT OF ADULT BASIC SKILLS:

LESSONS FROM CANADA

T. Scott Murray
Statistics Canada

Abstract

The measurement of adult literacy has usually relied on either self-reports or educational attainment as proxies for proficiency over a broad range of skills. While useful in heightening awareness of the nature and extent of the literacy "problem" in each country, such indirect measures suffer from a number of important methodological deficiencies that ultimately limit their usefulness for informing public policy. In the late 1980s and early 1990s, a number of countries fielded direct assessments of adult basic skills, and these have played a central role in defining a coherent public policy response to the literacy issue. This report describes the development and application of a related suite of "proxy" measures that both extend the usefulness of these assessment-based measures and offer interesting insights into the stability of the phenomenon in a cross-national perspective.

INTRODUCTION

The purpose of this report is to provide readers with an overview of proxy measures of adult basic skills. These measures are important in that they offer individual learners, educators, employers, and policymakers access to information from large-scale national assessments at a local level, information that would otherwise be unavailable to them. In the absence of local proxy measures, large-scale national assessments are either ignored by local decision makers, or national estimates are accepted as representative of local conditions without making the adjustments to the local demographic reality. As a result, decision making, and thus public policy, are poorly served. Proxy measures hold the potential to improve the social return on investment in large-scale assessments by improving local decision making. This report will draw heavily, but not exclusively, on Canadian experience with the development and application of proxy measures in the field of adult basic skills and, more particularly, adult literacy.

PROXY MEASURES DEFINED

Most college dictionaries define a *proxy measure* as something that has the authority or power to act for another. Thought of in a statistical sense, such authority or power can be defined technically, in quantifiable terms such as precision, bias, and reliability. Such a definition, however, reveals little of the potential of proxy measures for informing and possibly improving social policy.

Proxy measures are nothing new. In fact, proxy measurement is as old as measurement itself. Science, suggests W. Brian Arthur, an economist at Stanford University, "is about the creation of metaphor" (as quoted by Corcoran, 1992, p. 22). Proxy measures can be seen as a type of metaphor that allows order to be extracted from chaos. In the physical sciences, metaphor plays a minor role. Physical science deals in large measure with static relationships, which, once discovered, do not need to be reconfirmed through repeated measurement. Once discovered, one does not, for example, need to rediscover the freezing point of water. In the social sciences, metaphor plays a more central role. One continually confronts dynamic processes that are characterized by rapid and often unpredictable change. Where these changes have social or economic consequences, one is compelled to measure the phenomena repeatedly. Such measurement can be expensive and technically demanding. Through their close statistical association with the underlying phenomena of interest, proxy measures can serve as low cost surrogates for explicit measurement, allowing one to defer expensive measurement. They can also provide insight into the processes that drive differences observed across both countries and population subgroups. This information in turn has implications for the legitimacy of standardized protocols for the measurement of adult literacy. This report describes how the introduction of a new form of

measurement, the direct assessment of adult literacy, has led to the development of a related suite of proxy measures that can achieve this result while rendering the data accessible to a whole new range of uses.

A BRIEF HISTORY OF PROXY MEASUREMENT IN ADULT LITERACY

Although it may not seem obvious, proxy measurement has a long history in the measurement of adult literacy. In the Middle Ages, the ability to sign one's name was taken as the mark of an educated individual. In the terminology employed in this report, signing one's name in the Middle Ages was a reliable proxy measure because of its strong association with the ability to perform a broad array of intellectual tasks associated with formal education, including the cognitive tasks implied in reading and writing. In statistical terms, the two phenomena were seen as being almost perfectly correlated. As a result, those individuals able to sign their names could be assigned a very high probability of being able to perform the designated tasks, whereas those unable to sign could be assigned an equally low probability of being able to perform the intellectual tasks in question. In the late 19th century, the ability to read and write in the classical languages, Latin and Greek, was similarly taken as an indicator of higher learning.

Over time, the measures used as proxies of adult literacy have evolved considerably, with two distinct measurement traditions emerging (cf., Canadian National Literacy Secretariat, 1988; United Nations Statistical Office, 1989). The first approach has relied on self-reports of proficiency, usually embedded in national censuses of population, to provide undifferentiated estimates of literates and illiterates (Satin, 1991). The second measurement approach employs educational attainment as a proxy for estimating both basic and functional literacy levels. This approach has been enabled by the advent of compulsory schooling in all regions of the world. In Canada, the number of adults with less than Grade 5 was used as a proxy indicator to distinguish between basic literates and illiterates. Similarly, the percentage of adults not completing Grade 9 was taken as an indicator distinguishing between functional literates and full literates.

THE USES OF DATA ON ADULT LITERACY

It is important to arrive at some understanding of who the intended users of proxy measures on adult literacy might be, and for what reasons they would use the data. Three different categories of users can be distinguished,

namely government, program delivery institutions, and individuals. The information needs of each category are described below.

GOVERNMENT

Simplifying considerably, the role of government can be defined as a system for allocating scarce resources among a plethora of competing demands. How this process is managed and what objectives are the most valued in the process varies a great deal from country to country. In confronting any issue, however, governments must find answers to a basic set of questions:

- What is the nature and extent of the “problem”? Are enough people affected to justify action? Is any particular segment of society particularly disadvantaged?
- What are the costs associated with “correcting the problem”?
- What are the risks associated with inaction, expressed in terms of individual and collective prosperity and security?
- Where are the resources best invested? Which investments will provide the maximal social benefit?
- What system should be used to establish differential need in the allocation of funds?

Hence, it appears that government authorities need objective data to inform the decision-making process. It is important to note at this point that governments are obliged to make decisions regardless of the availability or quality of related objective information. For example, governments will decide if literacy is to be a priority or not. They will make choices between the relative needs of different groups based on whatever information they have at their disposal at the time the decision needs to be taken, irrespective of its quality or fitness for use.

PROGRAM DELIVERY

Although the structure, extent, and institutional locus of agencies and programs offering adult literacy instruction vary considerably from country to country, each country has a group of societal actors whose objective it is to deliver remedial literacy education to adults. These actors require information to serve three distinct needs. First, they require an assessment tool to diagnose the needs of individual students. Equipped with this information, curriculum and technique can be adapted to achieve maximum effect. Second, they require information to guide their own program and product development. Thought of as an industrial sector, the market for literacy goods and services is like any other market, constantly evolving as consumer demand and technology change. Hence, the literacy delivery organizations need to understand the needs of the consumers, whom they are currently serving and whom they could potentially serve, and what implications entry into new markets might have for product development. Finally, literacy delivery organizations often need information to justify government funding. Despite the market analogy used above, the market for literacy goods and services is an imperfect one, relying largely on government for revenue rather than on sales to consumers. As a result, literacy delivery organizations need data to attract money from funders, particularly government funders.

INDIVIDUALS

Individuals require information for two purposes. First, they have a requirement for information to define their own need for remedial training. Literacy is a phenomenon with social and economic consequences for individuals. As a result, it is in the interests of individuals to assess their own situation, to judge if a personal investment in remedial training will enhance their economic prospects or quality of life. Objective information on the likely returns to such an investment can complement anecdotal information gleaned from personal experience that individuals will bring to the decision. Second, individuals, as members of a democratic society, are obliged to come to an understanding of the relative importance of literacy compared to other social issues. Literacy is one of a multitude of social and economic issues competing for the attention of governments. To attract and hold the attention of government requires the implicit support of the electorate. Data on the nature, extent, and import of literacy serves the important purpose of marshaling societal awareness and maintaining the political will to justify continuing investment.

ADULT LITERACY DEFINED

In order to appreciate the development and application of proxy measures in the domain of adult literacy, it is important to convey the broad conceptual and definitional elements underlying the notion as it has been operationalized in Canada and the United States.

A variety of competing definitions have seen various degrees of use in recent years. These definitions can be roughly classified as being of two types—those that define literacy as a set of elementary reading and writing skills termed *basic literacy* and those that invoke a broader skill set termed *functional literacy*. In 1978, UNESCO offered the following definition for basic literacy: “a person is literate who can with understanding both read and write a short simple statement on his everyday life” (as quoted in Thomas, 1983). This definition naturally leads to a dichotomy wherein individuals are labeled either literate or illiterate.

In contrast, functional literacy encompasses a mastery of the knowledge and skills that an individual needs in order to participate fully in all aspects of society. Functional literacy includes the decoding skills involved in reading, writing, and counting, and also extends to other cognitive skills associated with decision making in society. UNESCO offered a second definition of literacy in 1978, which embodied a functional conception of literacy:

A person is functionally literate who can engage in all those activities in which literacy is required for effective functioning of his/her group and community and also for enabling him/her to continue to use reading, writing and calculation for his/her own and the community's development. (as quoted in Satin, 1991, p. 58)

Research in Canada and the United States suggests that the skill levels implied in a functional definition of literacy form a continuum, with each level becoming more complex and demanding, but each requiring both decoding and decision skills related to specific contexts of functioning. Rather paradoxically, there is also some evidence to suggest that the higher an individual's skill level, the more generalizable and transferable his or her skills become, allowing the individual to deal with unfamiliar tasks in unfamiliar contexts.

Lastly, the recent Canadian survey of the distribution of literacy skills in the adult population incorporated the following definition of functional literacy: "the information processing skills necessary to use printed material commonly encountered at work, at home, and in the community" (Murray, Kelly, Satin, & Montigny, 1980).

CRITERIA FOR JUDGING THE UTILITY OF PROXY MEASURES

In order to judge the utility of the prevailing approaches to measurement, it is important to have access to a framework that renders explicit the criteria for evaluation. This report adopts a highly modified version of the framework proposed by Goedegebuure and Overgaag (1991), two researchers at the Netherlands' Central Bureau of Statistics, for this purpose. The Dutch typology classifies the use of official statistics according to three dimensions: category of user, objectives of use, and methods of use.

For the purpose of this report, the above classification can be modified in three ways. First, the dimension of methods of use is not discussed, since it has more to do with the technical capabilities of the users and the kind of data products that are made available to them, than the fitness for use of the proxy measures themselves. In modern societies, virtually all users can avail themselves of the tools necessary to enable the entire range of analyses implied in the Dutch classification. Second, the different types of users set out in the classification are ignored. Different groups of users differ more in their tastes for various types of data products and services than they do in the demands placed on the actual data sources they use. Finally, the dimension of objectives of use has been collapsed. In their classification, Goedegebuure and Overgaag (1991) subdivide objectives of use into four categories—knowledge, preparation for choice and action, routinization of choice and action, and evaluation of choice and action—arguing that the value of any statistical information can only be judged in terms of its intended use. While on the face of it this is true, the underlying facts that distinguish these categories are the statistical attributes that the data must implicitly have to support valid use.

Of these, knowledge is perhaps the least demanding of the data, since scientific insight may be gleaned from as little as a single case and generalized upward to the population as a whole. Often, however, generalizations flowing from anecdote or intuitive leaps can only be validated through the application of rigorous scientific methodology, the purpose of which is to reduce the possibility that alternate hypotheses might explain the same phenomenon.

In the Dutch classification, preparation for choice and action, routinization of choice and action, and evaluation of choice and action are presented as distinct uses. One might conclude, therefore, that each use places unique demands in terms of statistical attributes. In reality, however, these uses differ little in the type or quality of data required for good decision making. Thus, they can be thought of as statistical equivalents if applied at the same level.

What does distinguish these uses is the level at which decisions flowing from use are made. For example, if one is making "high-stakes" judgments related to an individual's level of performance on an adult literacy scale, one needs to be far more certain that the data are reliable than if one is diagnosing the performance of entire populations (Mullis & Owen, 1994). By inference, the former application demands far more precise and unbiased estimates than the latter. A distilled framework for judging the utility of proxy measures might be proposed, a framework that incorporates two elements: first, the use to which the measure will be put, where use is further divided into knowledge acquisition and preparation for choice and action; and second, the size of the population at which the data will be applied, where the size is defined as either the individual level or some grouping of individuals up to and including the entire population.

For each cell in the resulting four-cell matrix, one can weigh the risk of being misled by the data against the cost of acquiring the information for any given application. The basic thesis of this report is that proxy measures can, with little increase in contingent risk and at marginal cost, deliver useful information to levels that cannot be estimated directly from the data. Implicit in this argument is the notion that any given measure may serve a variety of uses. It is inevitable, therefore, that the same measure will not meet all uses equally well. For some uses, a particular measure might be judged ideal, whereas for others it might be deemed wholly inadequate. The next section provides a description of problems with prevailing methods, in particular as they appear in light of the demand for information about adult literacy.

PROBLEMS WITH PREVAILING APPROACHES TO MEASUREMENT

Both traditions of measurement, those based on self-reports and those based on educational attainment, have been heavily criticized lately as inadequate. These criticisms have been of two sorts. The first suggests that a common definition of literacy is not feasible or desirable. Many writers in the fields of adult education and literacy contend that any rooting of the concept of functional literacy in context necessarily leads to pragmatic value choices over which skills are defined functionally necessary and whose interests and advancement are ultimately served. At the extreme of this position lie a group of people who profess that measurement is, in and of itself, an offensive act that in some way diminishes both the humanity of those being measured and their control over their destiny. To adopt such a position is, in the author's view at least, to forego the intrinsic power of scientific knowledge to

improve the world. As mentioned above, science is about the discovery of metaphor and aims at describing underlying processes as a means of understanding cause and effect, and making generalizations that do not offend reality. Armed with such information, individuals and governments can make informed choices with some knowledge of the outcomes that they will precipitate. In short, without science and the reliable information it brings, one is forced to make decisions at the individual or national level that are based on anecdotal evidence or personal prejudice.

The second line of criticism has been directed at the statistical properties of the measures themselves. In the words of Satin (1991),

the interpretation of statistics on literacy and the ability to reconcile international figures is affected by the uneven levels of data quality associated with operational instructions. The clarity of operational instructions and compliance with them affects the level of misunderstanding and misclassification which can result. The instructions are, as well, generally prone to a large degree of subjectivity in their application. This can be exacerbated depending on the extent to which proxy reporting is allowed on behalf of individuals within households. The social desirability of being considered illiterate adds to the problem of misclassification. (p. 58)

In the language of the statistician, such words amount to a damning indictment indeed, the implication being that the estimates so derived are so full of measurement error as to make them useless. It is reasonable to argue, however, that these criticisms are unjustified *given the use to which they were designed to be put*.

Jones (in press), in a defense of measurement in adult literacy, observes that while contests among ultimate values undoubtedly form an important component of the anchoring of any concept into context, too close an adherence to this view would deny the possibility of societal standards for both achievement and development. Such standards certainly do exist. Within market societies, such as Canada, it is frequently argued and easily demonstrated that rather harsh conditions of life prevail for the undereducated. Literacy skills in the population, therefore, are a principal element in both personal and collective economic prosperity. In the past few years, many OECD countries have had to face a new fact—their historic economic well-being is in some measure due to literacy and continued renewal and well-being may critically depend on literacy. As long as the national government assumes a central role in ensuring economic prosperity and equity over the long term, the imposition of certain definitions is warranted for certain purposes.

With respect to the statistical challenge, the answer lies elsewhere. Estimates based on educational attainment were primarily intended to demonstrate the beneficial impact that the introduction of compulsory education had on the magnitude of the literacy “problem” in the developing countries of the Third World. The extension of schooling entailed the investment of significant social and economic resources on the part of these nations, and the data served the important purpose of marshaling societal awareness and maintaining the political will to justify continuing investment. For this use, which can be broadly categorized under the rubric of *preparation for choice and action*,

imprecise and biased estimates were good enough to meet the needs of government and the literacy delivery organizations.

For the industrialized countries, however, the adequacy of these measures was to become increasingly less clear. By the mid-1970s, using educational attainment data as the standard, mass education should have virtually eliminated "basic" illiteracy. Yet governments were continually confronted with evidence that basic literacy remained a pernicious problem. Evidence mounted that educational attainment was, in fact, a poor proxy of literacy, falsely labeling both literates and illiterates (cf. Wagner, 1992). Strong positive, but certainly less than perfect, correlations exist between grade attainment and real world skills. Some persons with quite high grade levels are not fully functional, and conversely, some individuals with only modest education are highly literate (Neice & Adsett, 1990). This finding challenged the standard conception of how literacy is acquired, the so-called *static model*, wherein a known quantum of skills is accumulated in a burst of childhood schooling to be retained indefinitely. What seemed to emerge was a complex relationship wherein some adults found ways outside the system of formal education to acquire the requisite reading skills, whereas others apparently found ways to lose the skills they once had. Such a dynamic notion of literacy skill acquisition and loss offers a much more realistic picture of the complexity of the issue.

In addition, the governments of OECD countries began to appreciate the policy implications of the context-driven, multilevel, functional measures of adult literacy (cf. Van der Kamp, in press). For the first time, they had a conceptual framework that allowed for a literacy skill deficit to be defined in terms of a mismatch in supply and demand across a differentiated continuum of skills. In the next section of this report, it will be seen that this realization was to have a profound impact on both policy formation and measurement.

THE ADVENT OF DIRECT SKILL ASSESSMENTS IN NORTH AMERICA

The answer to the problem of measurement and interpretation discussed in the preceding section is the advent of the direct approach to skill assessment. An abbreviated review of the key recent North American studies is reported by Neice and Adsett (1990) and Venezky (1992). These authors consider these direct assessments as a fusion of educational testing with survey research technology. The resulting assessments involve surveys in which large, statistically representative samples of adults are actually tested to determine their ability to deal with a variety of printed materials drawn from everyday experience, be it at home, at work, or in the community. The Canadian study, called the Survey of Literacy Skills Used in Daily Activities (LSUDA), tested the official language reading, writing, and numeracy abilities of 9,500 adults using the approach developed for the United States National Assessment of Educational Progress (NAEP) study of young

18

adults. The primary statistical output of LSUDA is a population parameter based on scaled test scores for reading, writing, and numeracy. Generated using a two-parameter logistic regression in an Item Response Theory (IRT) framework, these scores are subsequently collapsed into the following skill levels to aid in interpretation (see also Kirsch & Jungeblut, in press; Mosenthal, in press):

- Level 1 Canadians at this level have difficulty dealing with printed materials. They most likely identify themselves as people who cannot read.
- Level 2 Canadians at this level can use printed materials only for limited purposes such as finding a familiar word in a simple text. They would likely recognize themselves as having difficulties with common reading materials.
- Level 3 Canadians at this level can use reading materials in a variety of situations provided the material is simple, clearly laid out, and the tasks involved are not too complex. While these people generally do not see themselves as having major reading difficulties, they tend to avoid situations requiring reading.
- Level 4 Canadians at this level meet most everyday reading demands. This is a large and diverse group that exhibits a wide range of reading skills.

Each of these levels in effect defines a different aspect of the literacy "problem," each with its own set of unique implications for remedial policy and programs.

It is important to note that direct assessments themselves provide only proxy measures, albeit ones that are much more directly associated with the fundamental phenomena of interest. There exists the possibility, however, that tests designed to measure one trait, such as LSUDA's measurement of literacy, actually measure multiple traits. Fortunately, although there is some debate as to which techniques are best, the psychometricians tool kit includes statistical tools to test for such "dimensionality" (Gessaroli, 1992). Generally, the direct assessments have passed these tests with flying colors, even though some issues of validity remain to be addressed.

It is useful, therefore, to examine the uses to which assessments such as LSUDA can be put. The LSUDA assessment was designed to provide knowledge, by shedding light on the relationships between measured proficiency and a variety of explanatory variables such as education, labor market activity, and reading behaviors. At a very basic level, it was also intended to test the coherence of the conception of literacy, developed by Kirsch and Jungeblut, and Mosenthal (Kirsch & Jungeblut, in press; Mosenthal, in press), upon which the assessment was based (Kirsch & Mosenthal, 1990; Mosenthal & Kirsch, 1989). The use of the assessment data to generate new knowledge was satisfied primarily through the provision of a microdata file, including individual item scores and scaled test results, to academics and literacy researchers.

A second major objective of the LSUDA study can be thought of as serving the needs of those institutions involved in planning and delivering remedial

literacy programming in Canada. This use focused on estimating the magnitude of the literacy "problem" for various geographies and clientele, and then determining the characteristics of those needing literacy services. The latter aim was particularly important in the Canadian context. The traditional literacy providers found themselves facing a shortage of their traditional clientele—the illiterates. Survival alone dictated that they consider reaching out to other clientele farther up the literacy skill continuum. Those institutions deciding to serve new markets were forced to develop new products and services and implement new delivery mechanisms for their programs. For example, those institutions choosing to serve Levels 2 and 3 in the Canadian classification were obliged to forge new partnerships with employers and unions to deliver their programs since most of the individuals at these levels were employed.

The same data were also used in the determination of funding levels and allocations by the federal, provincial, and municipal governments. In the modified Dutch classification reviewed above, such uses are classified as preparation for choice and action.

The LSUDA has also had a profound effect on individual citizens in Canada, largely by introducing a new paradigm of literacy to the public consciousness and debate. Until recently, the policy debate relating to adult literacy in Canada has invoked a conception of literacy that divided the world into "literate" and "illiterate." While useful in helping national governments define the extent of the literacy "problem," the definition of literacy in such absolute terms has limited the discourse in a number of important ways. First, literacy has come to be treated as a pathological condition, as a disease that afflicts an unfortunate few. In severe cases, the victims are even blamed for infecting their offspring. As a consequence, illiterates have come to be viewed as flawed characters, devoid of skills. The resulting stigma often drove away those most in need of remedial education (Hautcoeur, 1990). A second unfortunate consequence of this definition has been to suppress any debate about the adequacy of the skills of those judged to be "literate." There is increasing evidence that literacy is a relative concept that can only be judged in terms of the demands of society. In times when societal demands are growing, such as today, even "literate" may need remedial education.

By comparison, the conception of literacy embedded in the direct assessments conducted in North America can be viewed as liberating in a number of respects. The most important notion that is introduced is that literacy is a continuum of skills in which everyone has some level of proficiency. Such a conception opens the way to a debate that is centered around the match between individual skills and societal demands rather than around individual deficiencies. It also frees the issue of pejorative labels, which may disrupt debate and participation in remedial education and training.

To date, the literacy assessment protocol used in LSUDA has not been applied to the diagnosis of individual performance for the purposes of defining remedial training. Given that the measurement technology is rooted in a robust understanding of adult reading, this would represent a natural extension of the research tool in a way that can be thought of as both routinization and evaluation at the level of the individual. The Educational Testing Service is, in fact, in the final stages of developing a computer-

driven group-instruction system that incorporates this approach to the diagnosis and placement of individual learners.

It is worth noting that such coherence of approach to the problem of measurement at all levels offers great benefits to the entire spectrum of users. At a minimum, it enables communication, since everyone will be using a common conceptual framework and lexicon to describe the issue.

Apart from the Canadian and U.S. assessments already noted in this report, a number of conceptually related studies have been fielded in the Netherlands (Doets, Groen, Huisman, & Neuvel, 1991), Spain (Flécha et al., 1993), France (DEP/CREDOC, 1992), Australia (Wickert, 1990), and Romania. For the most part, these studies have been designed to meet the same set of needs as LSUDA.

IMPLICATIONS OF DIRECT ASSESSMENTS FOR PROXY MEASURES OF ADULT LITERACY

From the previous section, it is apparent that direct assessments have transformed the policy debate in a number of positive ways. Ideally, one would like to have similar estimates available for every level of decision making. Unfortunately, the cost of providing these data would be impossibly high, well beyond the means of the societal actors involved to bear. For a small investment, however, national level data from direct assessments can yield information to improve existing proxies and to suggest new measures for the local level. The Canadian experience offers some interesting insight into these possibilities.

CORRECTION FACTORS FOR ESTIMATES OF EDUCATIONAL ATTAINMENT-BASED PROXY MEASURES

The first example to be highlighted can be traced to the seminal work of Neice and Adsett (1990), in which they use multivariate techniques to explore the statistical relationships between performance, as measured by the LSUDA test, and a variety of demographic and other variables collected by LSUDA. Their analyses reveal a number of interesting things. First, they document the actual extent and nature of the imperfect relationship between tested ability and educational attainment. This finding opened the way to the calculation of "correction" factors for improving existing proxy measures of adult literacy levels. Without correction, the literacy proxies based on educational attainment reveal the picture of the literacy "problem" in Canada shown in Table 1 (see Appendix). As illustrated in Table 2 (see Appendix), the distribution of literacy skills obtained through LSUDA appears to be quite similar to that derived through the UNESCO classification. As Table 3 (see Appendix) vividly shows, however, the apparent congruence of these measures obscures the fact that

traditional proxy measures based on the educational attainment both over- and underestimate the actual performance of significant proportions of the population when data from direct assessment are used to put "misclassified" individuals in their correct level.

When applied to educational attainment measures derived from the Census of Population or from ongoing household surveys such as the Labor Force Surveys, which are conducted by most OECD countries, application of "correction" factors implied in the above data would provide more reliable proxy estimates of adult literacy than those based on "uncorrected" educational attainment measures. Since most countries conduct a decennial census of population, even relatively small direct assessments can serve to provide sufficient information value to correct the profile of educational attainment, thereby providing more reliable estimates of the magnitude of the literacy "problem" for policy relevant subpopulations. A conservative approach to calculating "corrected" estimates would limit the calculation of correction factors to those subpopulations for which sufficient sample was collected in the assessment to support direct estimation of their skill profile. As explained below, however, to restrict the estimation in this way is to ignore much of the information value that can be gleaned from survey assessment data.

The application of correction factors to adjust educational attainment data and hence obtain current estimates of literacy ability can be used to forecast the supply of skill and the demand for it if one is willing to live with the assumption that the observed relationships between tested performance and educational attainment are stable over time. This being the case, one can apply the same factors to projections of educational attainment by age cohort, geography, or other demographic variables generated from censuses or surveys to yield a glimpse at how the problem might evolve in the near term. As noted above, this type of information plays a central role in the policy formation deliberations of a variety of social actors in the literacy domain, social actors who are obliged to make decisions even in the absence of such data.

IDENTIFICATION OF NEW PROXY MEASURES WHICH COULD BE COLLECTED

The second example of how direct assessment data have been used in Canada also stems from the work of Neice and Adsett (1990). In addition to the correlation analysis referred to above, they used a variety of multivariate regression techniques to examine the statistical relationship between tested performance, self-assessed literacy ability, reported frequency of reading, and a variety of other individual characteristics and behaviors. For example, Neice and Adsett (1990) discovered that the frequency of reading books, collected as a variable on the assessment background questionnaire, was a good indicator of literacy skill to be found in the data, accounting for up to 46% of the observed variance in level. If this variable were collected on an occasional basis and paired with information on educational attainment, the resultant information used could be used to derive current estimates and projections of the literacy skill profile in the absence of concurrent testing. It is expected that these estimates would reflect up to 90% of the observed variance in literacy level, providing relatively little room for error that might distort the data. Such an approach is attractive since the cost of collecting

information on reading behaviors is magnitudes lower than those associated with generating new test results. Again, a critical assumption is that the relationships observed at the base period remain stable over the reference period to which the estimates pertain.

PROXY SKILLS FOR SMALL AREAS

A third Canadian example of direct assessment data spawning new and/or improved proxy measures is provided by Murray and Shillington (1991). In this study, the statistical tools used were similar to those employed by Neice and Adsett, but the object of analysis was somewhat different. The objective of the exercise was to build a predictive model of literacy ability. By making a conscious decision to restrict the predictive variables to those that were available on the Canadian Census of Population, Murray and Shillington (1991) were able to devise a robust estimation methodology for producing literacy skill profiles for small geographic areas, areas that are far too small for even the relatively large Canadian assessment to estimate directly. For each cell in a cross-classification of the predictive variables, the model estimates the proportion of individuals that will be found at each level in the literacy skill profile. If the model performs well, it will account for a sufficiently high proportion of the variance observed between cells in the "demographic matrix" to apply the resultant probabilities to fine-grained census tabulations of the same variables. The result is what can be thought of demographically corrected proxy skill profiles for units such as federal electoral districts and census subdivisions. In Canada, such units are important politically, since they represent the level at which most of the funding allocation decisions are taken and remedial programming is delivered. Without such a product users are forced to live with the provincial or national estimates provided by the direct assessment or to make up their own "adjusted" estimates based on intuition. An example of one of these small area literacy estimates is provided below.

A total of 32 models were tested, and the results show that virtually every one of the models accounted for a significant proportion of the observed variance in literacy level. One model captured fully 80% of the observed variance, whereas the final model, selected for its goodness of fit, still explained roughly 75% of this variance, leaving very little room for serious error.*

A variant of this analyses has been performed on the U. S. data by Kirsch and Jungeblut (1993). Overall, they were able to account for much less of the observed variance than their Canadian counterparts, capturing between 52% and 78% depending on the population of study and the variables used. They are led to conclude that proxy measures of this type are of limited utility except when applied to large, heterogeneous populations, arguing that the debate has moved beyond trying to estimate "how many." Yet in Canada, at least, government's need for objective information to inform the allocation of scarce financial resources among competing geographic areas remains very much a question of "how many." The question that must be confronted is "Do the estimates allow decision makers to make more informed choices since they are obliged to make these choices even in the absence of any objective information?" A strong case can be made that these estimates are much superior to what they replace (i.e., no information or information that has not been corrected to reflect differences in

* Details of the models and the survey are available from the author.

the characteristics of the local population), which are known to have an impact on the distribution of literacy ability.

A second, less apparent, reason for investing resources in the statistical analysis needed to support the production of small area estimates can be found in the wealth of information that the regression process yields in and of itself. A comparison of the Canadian and U. S. results reveals some interesting differences in the relative importance of predictive variables, differences that might have both policy and academic import. One also needs to confront the possibility that the observed differences are attributable to differences in the measurement of the background variables themselves rather than in the underlying relationships. Finally, it can be argued that the inability to find stable patterns across population subgroups and countries can mean one of two things, either the "right" variables were not included in the background questionnaires, or literacy, as a social phenomenon, defies generalization.

IMPLICATIONS OF PROXY MEASURES FOR DIRECT ASSESSMENTS

The availability of new and improved proxy measures afforded by direct assessments can also affect the design and output of the direct assessments themselves. The first of these effects to be considered concerns the availability of more reliable proxy measures that will allow researchers to extend the length of time between assessments. Historically, at least, most change in performance over time has been attributable to differences in performance between "incoming" and "outgoing" age cohorts. Without meaning to denigrate or discount the efficacy of adult remedial training, relatively little of the observed differences can be attributed to adult learning. One can now, with a modest analytic investment, delay the conduct of a new assessment until there is evidence that the historical patterns have been disrupted in some way. As Neice and Adsett (1990) have noted, direct assessment is extraordinarily expensive and relies on narrow technical expertise that is in short supply. Anything that can obviate cost and reduce the overall need for scarce technical resources is therefore a welcome development.

The second major effect that must be considered is that the availability of reliable proxy measures will force the designers of direct assessments to reconsider the basic objectives of such studies. These studies are currently designed to generate aggregate estimates of the *numbers* of individuals at a given level of literacy proficiency. Experience has shown that one needs a yield of roughly 600 cases to support an estimate for any given group. As a result, total sample sizes have tended to be large. For example, charged with producing univariate estimates for each of Canada's ten provinces and bivariate estimates for five regions, LSUDA demanded a sample of some 9,500 cases. Similarly, the recently completed National Adult Literacy

Survey (NALS) in the United States demanded almost 20,000 respondents. Together with the costs inherent in testing an adult population, these sample sizes contribute to the extraordinary costs of conducting direct assessments.

Were study designers willing, however, to view their objectives solely in terms of supporting improved proxy measures by exposing *relationships* between variables (rather than producing point estimates) one could reduce sample sizes by at least an order of magnitude. These reductions would be achieved in two ways. First, the multivariate techniques described above are far less demanding in terms of sample size, being driven by the strength of the observed relationships rather than by the distributional assumptions that drive the traditional sample size calculations. About 50 cases are often enough to discern statistically meaningful differences between groups. Second, this approach offers the opportunity to optimize the allocation of the available sample to generate the best signal-to-noise ratio for any given sample size. This effectively involves reducing the sample allocated to population subgroups where the predictive power of the observed relationships is strong, so that the sample can be augmented in strata where the relationships are less predictable. On balance, a case can be made that the impact on "fitness for use" would be little changed.

The third major aspect to consider concerns the multivariate analysis needed to support the generation of these new proxy measures. Such analysis can inform both the design of future assessments and the content and structure of the social policy response. However, the failure of the model to account for a sufficiently high proportion of the observed variance suggests two lessons for the design of future assessments. First, the predictive variables chosen for inclusion have not been measured well. Both sampling and nonsampling error can contribute to measurement error in these variables. This fact argues strongly for the involvement of statistical agencies in the design and collection of assessment data, since they alone have the resources, interest, and mandate to contain such error. Second, the current assessments have excluded important explanatory variables that should be incorporated into future assessments to capture unobserved heterogeneity. This argues strongly for the involvement of literacy researchers in the formulation and design of future literacy assessments since they are the most likely source of information on the "missing" determinants of literacy skill.

As for the structure and content of the policy response to the prevailing distribution of literacy skill, the multivariate analyses also carry an important message. If the literacy outcomes, as defined by levels of proficiency, can, in fact, be predicted reliably with a small number of demographic and socioeconomic variables, macro-level remedial policies have a high probability of being successful. If, on the other hand, one requires large numbers of variables to reliably predict the observed outcomes, or if one were unable to explain a significant proportion of the observed variance in the outcome, then micro-level solutions, which leave control and decision making at the individual or local level, would be favored. It is the tension between these two views that underlies much of the debate over the legitimate role of empirical assessment in the field of literacy, a tension that will not be resolved unless and until researchers demonstrate that their product is both valid and informative rather than oppressive.

CONCLUSION

The above discussion holds a number of important messages for those concerned with the construction of international indicators of adult education. The first notion that bears repeating is a caution that current proxy measures, such as educational attainment and years of schooling, are either so fraught with measurement error, or so uncorrelated to the actual performance of adults, as to be useless for informing public policy. The second lesson of note is that direct assessments can, apart from generating estimates that are more reliable than existing proxies, also be used to recalibrate these same proxy measures to provide much more reliable estimates for use in the subnational context.

The usefulness of these techniques internationally is currently constrained. International comparison would demand administration of a common, standardized test in all countries to be compared. It is only in this way that national proxy measures could be adjusted to reflect a common denominator. Organizations involved with school-based assessments, such as the International Association for the Evaluation of Educational Achievement (IEA), have 30 years of experience in the design and fielding of such studies upon which to draw.

With respect to measures of adult education, the first tentative steps are being taken in the form of the International Adult Literacy Survey (IALS). Jointly developed by Statistics Canada and the Educational Testing Service of Princeton, New Jersey, IALS will simultaneously administer a common test of adult literacy in Canada, the United States, France, Germany, the Netherlands, Switzerland, Ireland, Sweden, and Poland (cf. Statistics Canada, 1992). Having involved assessment experts, statistical agencies, and policymakers in the design process, it is hoped that this study will yield reliable, policy-relevant data that will support statistically legitimate international comparison.

The final conclusion is that much remains to be done. This report has purposefully restricted itself to a discussion of the literacy dimension of adult basic skills, specifically reading literacy and numeracy. It is clear, however, that these skills are just two of a much broader skill set that has social and economic import. For example, both problem-solving ability and communication skills have been noted as important factors in defining worker productivity and adaptability (OECD, 1991). Although a fairly rich measurement literature exists for these abilities, most of this work has been involved with extensive diagnostic testing of individuals in school or in the workplace. Much collective effort needs to be focused on adapting these measures for use in the household survey context for the purposes of making inferences to the population level. This is the only way in which a differentiated, statistically reliable set of national and international indicators can be created in the field of adult education.

REFERENCES

- Canadian National Literacy Secretariat. (1988). *Definitions, estimates, and profiles of literacy and illiteracy* (Working Paper #1). Ottawa: Department of the Secretary of State for Canada
- Corcoran, E. (1992, October). The edge of chaos. *Scientific American*, 17-22.
- DEP/CREDOC. (1992, April). *Regards sur des adults en recherché d'insertion professionnelle* (Document de Travail 403). Paris: Ministère de l'éducation nationale et de la culture, direction de l'évaluation et de la prospective and Ministère du Travail, de l'emploi de de la formation professionnelle. Groupe permanent de lutte contre l'illettrisme.
- Doets, C., Groen, P., Huisman, T., & Neuvel, J. (1991). *Functional illiteracy in the Netherlands*. Amersfoort: SVE, CIP-Gegevens Koninklijke Bibliotheek, Netherlands Study and Development Centre for Adult Education.
- Flécha, R., Lopez Palma, F., Recasens, F., et al. (1993). *Analfabetismo funcional a Catalunya* (Working paper). Barcelona, Spain: Generalitat de Catalunya, Direccio General de Formacio d "adults."
- Gessaroli, M. E. (1992, December). *A psychometric assessment of the equivalence of translated versions of the survey of literacy skills used in daily activities* (Technical analysis). Ottawa: Statistics Canada.
- Goedegebuure, R. V., & Overgaag, P. H. A. (1991). A classification of types of use of official statistics. *Statistical Journal of the United Nations, ECE* 8, 217-232.
- Hautcoeur, J. P. (1990). *From understanding of illiteracy as a pathologic problem towards a positive consideration of the competencies of people that society call illiterates*. Hamburg: UNESCO-UIE.
- Kirsch, I., & Jungeblut, A. (1993, Spring). *Scaling literacy assessment data and exploring variation in performance: An initial investigation*. Princeton, NJ: Educational Testing Service.
- Kirsch, I., Jungeblut, A. (in press). *Using large-scale assessment results to identify and evaluate generalizable indicators of literacy* (Technical report). Philadelphia: University of Pennsylvania, National Center on Adult Literacy.
- Kirsch, I., & Mosenthal, P. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, Winter, XXV/1.
- Jones, S. (in press). Measuring adult basic skills. In A. Tuijnman, I. Kirsch, & D. Wagner. *Adult basic skills: Innovations in measurement and policy analysis*. Cresskill, NJ: Hampton Press.
- Mosenthal, P. (in press). Using large-scale assessments to set educational policy: An agenda-analytic approach. In A. Tuijnman, I. Kirsch, & D. Wagner, *Adult basic skills: Innovations in measurement and policy analysis*. Cresskill, NJ: Hampton Press.
- Mosenthal, P. & Kirsch, I. (1989, October). Understanding Documents: Lists: The building blocks of documents. *Journal of Reading*, 58.
- Mullis, I. V. S., & Owen, E. H. (1994). The monitoring of cognitive outcomes. In A. C. Tuijnman & T. N. Postlethwaite (Eds.), *Monitoring the standards of education* (pp. 133-150). Oxford: Elsevier Science and Pergamon Press.
- Murray, T. S., Kelly, K., Satin, A., & Montigny, G. (1988). *A national literacy skill assessment planning report*. Ottawa: Statistics Canada.
- Murray, T. S., & Shillington, R. (1991, September). *Estimates of literacy skill for small areas*. Ottawa: Statistics Canada.

- Neice, D., & Adsett, M. (1990, September). *Direct versus proxy measures of adult literacy: A preliminary re-examination* (Working Paper #2). Ottawa: Social Trends Analysis Directorate, Department of the Secretary of State of Canada.
- Organisation for Economic Co-operation and Development (OECD). (1991). *Adult illiteracy and economic performance*. Paris: Centre for Educational Research and Innovation.
- Satin, A. (1991). An international review of the concepts, definitions and measurement approaches underlying literacy statistics. In *Adult literacy in Canada: Results of a national study*. (Statistics Canada Catalog #89-525, September)
- Statistics Canada. (1992, December). *An international assessment of adult literacy: A proposal*. Ottawa: Author.
- Thomas, A. M. (1983). *Adult literacy in Canada—A challenge*. Ottawa: Canadian Commission for UNESCO.
- United Nations Statistical Office. (1989). *Measuring through household surveys: A technical study on literacy assessment and related education topics through household surveys*. National Household Survey Capability Programme, United Nations Department of Technical Cooperation for Development, and United Nations Statistical Office DP/UN/INT-88-XOI/IOE.
- Van der Kamp, M. (in press). European traditions in literacy research and measurement. In A. Tuijnman, I. Kirsch, & D. Wagner, *Adult basic skills: Innovations in measurement and policy analysis*. Cresskill, NJ: Hampton Press.
- Venezky, D. (1992). *Matching testing literacy with social policy: What are the alternatives?* Policy Brief No. PB92-01. Philadelphia: University of Pennsylvania, National Center on Adult Literacy.
- Wagner, D. (1992). *Literacy: Developing the future*. Paris: UNESCO.
- Wickert, R. (1989). *No single measure*. Sidney, Australia: Commonwealth Department of Employment, Education and Training.

APPENDIX A: TABLES

Table 1	<i>Number and Percentage of Adults in Canada by UNESCO Literacy Level</i>	A-iii
Table 2	<i>Number and Percentage of Adults in Canada by LSUDA Literacy Level</i>	A-iv
Table 3	<i>Percentage of Adult Population in Canada Whose Literacy Skill Level Is Misclassified by Educational Attainment</i>	A-v

Table 1

Number and Percentage of Adults in Canada by UNESCO Literacy Level

	Number estimated to be at level	Percentage
Illiterate (completed less than Grade 5)	78,339	0.4
Basic literate (completed Grade 5-8)	1,960,533	11.1
Functionally literate (completed Grade 9 or higher)	15,666,081	88.5

Note: From LSUDA, Statistics Canada, 1989.

Table 2

Number and Percentage of Adults in Canada by LSUDA Literacy Levels

	Number estimated to be at level	Percentages
Illiterate (completed less than Grade 5)	863,364	4.9
Basic literate (completed Grades 5-8)	1,698,433	9.6
Functionally literate (completed Grade 9 or higher)	15,143,156	85.5

Note: From LSUDA, Statistics Canada, 1989.

Table 3

Percentage of Adult Population in Canada Whose Literacy Skill Level Is Misclassified by Educational Attainment

Ability according to UNESCO Educational Attainment Proxy	Estimated number at level	Percentage under-classified according to tested ability	Percentage over-classified according to tested ability
Illiterate (< Grade 5)	21,589	31.1%	0%
Functionally literate (Grades 5-8)	1,646,939	53.8%	12.8%
Fully literate (> Grade 8)	1,285,368	20.2%	7.0%

Note: From LSUDA, Statistics Canada, 1989.

Note: < Grade 5 should, according to UNESCO, be at level zero or one, people at level two or better are under classified.

< Grades 5-8 should be in level two not in level one (over classified) and not at level three or four.

> Grade 8 should be in level four.