

## **Applied Developmental Assessment**

Presentation in the Donald L. Peters Symposium

April 23, 2004, University of Delaware

*Nancy A. Busch-Rossnagel, Fordham University*

Fordham University is a leader, particularly among psychology departments, in the relatively new field of applied developmental science. Many people have been associated with Fordham's leadership in this area. If I can take any credit for Fordham's work, I would trace my involvement to seeds that Don Peters planted more than 25 years ago at Penn State. Specifically, the minor in special education that resulted from my contact with Don as a PrePAIR (PREdoctoral Preparation in Applied Interdisciplinary Research) trainee required that I understand the process of individual assessment. Likewise, Don's course in program evaluation formed the basis for Fordham's applied developmental practicum. Most importantly, though, Don's sensitivity to the differences among individuals in our society, such as those with disabilities, led me to a career exploring the diversity of human development and the contexts in which that development occurs. Honoring Don, I would like to tell you today how applied development science approaches individual assessments and program evaluations so that they can be culturally-centered and thus reflective and respectful of the diversity of today's society.

Assessment has become a fact of life for today's children and their families. School testing grew after the 1983 report *A Nation at Risk* (National Commission on Excellence in Education, 1983) documented that Americans lagged behind students in other developed countries in school performance. The response was to articulate academic standards and then to create tests to assess if the standards were being met. By Spring, 2002, the result was Public Law 107-110, the

No Child Left Behind Act, that sets specific goals for performance on standardized tests and ties federal aid for education to the testing results.

The federal mandates represent a traditional “one size fits all” approach that can provide an assessment of minimum competencies, but this approach is ill suited to understanding the complexities of individual human development or of changes as a result of educational reforms. Work in applied developmental psychology suggests a different method, which I call *applied developmental assessment*. Characterized by a utilitarian, differential, process approach, applied development assessment shows a heightened sensitivity to validity, both in research design and measurement. The entire process requires collaboration between the assessor and the individual or group being assessed.

### **Utilitarian Approach**

First, to be “applied,” assessment should have **purpose and utility**. There are different reasons for assessment, most of them related to the well being of the person or group being assessed. One reason for assessment is screening, or to identify individuals at risk for poorer developmental outcomes. Screening also serves the purpose of reducing the number of individuals who receive more in-depth assessment.

Traditionally, this in-depth assessment provides information for diagnosis, a second reason for assessment. However, the concept of diagnosis is intimately tied to the medical model, which looks for evidence of a specific disorder. For the applied developmental psychologist, this second level of assessment is not limited to identifying a disorder, and diagnosis should be replaced by creation of a conceptual framework for the problem or issues. A conceptual framework should provide information on the child=s developmental strengths and vulnerabilities and on their

context to provide a foundation for understanding the problem or issues that brought the child to the assessment.

The last stage of applied developmental assessment is to apply this conceptual framework to the identification of an intervention strategy through program planning. Applied developmental interventions are usually designed as programs serving more than one individual and should help overcome disabilities, which are functional limitations associated with impairments, prevent handicaps, or the secondary problems resulting from limitations in social roles and relations, and optimize developmental pathways (Busch-Rossnagel, 1981). Effective interventions require evaluation at both the individual level (monitoring changes) and at the group level, so program evaluation is another reason for assessment.

Applied development science emphasizes that effects are bidirectional, for example children influence their parents, as well as parents influencing children, that individuals are both producers and products of their own development (Lerner & Busch-Rossnagel, 1981a). Consistent with this understanding of bidirectional effects, assessment has utility not only for the person or group being assessed, but also for the professional involved. Specifically, applied developmental assessment provides professionals with a scheme of normative development and information about the range of individual differences.

### **Differential Perspective**

In addition to being purposeful, applied developmental assessment should also be “developmental.” Developmental assessment has traditionally referred to normative, age-graded assessment (such as the establishment of major milestones). However, chronological age, with the assumption of linear development, may not be the best marker for time-related changes, especially after infancy (Wohlwill, 1973). Instead, applied developmental assessment should be

characterized by recognition of individual differences and by plasticity, that is, the potential for systematic change across the life-span (Lerner & Busch-Rossnagel, 1981b).

In statistical terms, this means forsaking models, such as MANOVAs, that take an incremental view of change and solely examine mean differences over time. An approach such as individual growth curve modeling examines both mean levels of functioning and rates of change and thus can provide information about interindividual differences and intraindividual changes (Francis, Schatschneider & Carlson, 2000).

### **Process Approach**

A third characteristic of applied developmental assessment is that it should be a process rather than being equated with “testing.” A test is an evaluative device or procedure in which a sample of an examinee’s behavior in a specified domain is obtained and subsequently scored using a standardized process (AERA, APA & NCME, 1999, p. 3). For a test, the behaviors and context are the means, but the desired result, the end product, is the score. Assessment is a larger process that may—or may not—include tests, but which should examine behavior in a variety of settings, the meaning of the behavior in terms of individual’s functioning, and the likely explanation for that functioning (the conceptual framework). In other words, the means become the desired ends.

The fact that assessment is not equated with testing does not mean that applied developmental assessment is free to ignore psychometric principles. The contrary is in fact true: applied developmental assessment should be characterized by additional attention to the psychometric qualities of tests, observations, etc. in different contexts. In this way we can discriminate bias from random error.

### **Design Validity**

One type of bias is the lack of cultural equivalence of measures, and applied developmental assessment is particularly suited to identify such bias because it emphasizes the influences of diversity on development. In examining the validity of developmental assessments for subcultural groups, researchers may be tempted to turn to the classic research designs, especially that of the experiment. In the experiment, independent variables are manipulated by the researchers to observe their effects. The true experiment also requires random assignment to the control versus the experimental group.

When applying the design of the experiment to subcultural groups, researchers often use culture as an independent variable in an attempt to understand behaviors. However, culture is a sociobiographical variable, which cannot be manipulated easily; in particular it cannot be randomly assigned to research participants. Thus, research using culture as an independent variable is really a “pseudo experiment” and will be limited in its ability to establish causality and explain behavior (Busch-Rossnagel, 1992).

Because such research using culture as an independent variable holds limited promise for establishing the why of behavior, much cross-cultural or subcultural work is conducted to examine the population generalizability of tests or results of previous research. This approach still uses culture as an independent variable, which renders interpretation of results problematic. Research on population generalizability can yield two results: significant differences between the two groups or no significant differences. The finding of no significant differences is essentially meaningless because it is an acceptance of the null hypothesis.

The finding of significant differences suggests the rejection of the null hypothesis of population generalizability, but this result may also be uninterpretable because culture or ethnicity is usually confounded with other variables, particularly socioeconomic status or educational

attainment. Consider as an example the results of a study by Luis Laosa (1980), who found differences between Chicana and Anglo mothers when observing their teaching behaviors. When he controlled for the differences in the educational levels of the mothers, the ethnic group difference disappeared. Thus, when we use pseudo experiments to look at cultural groups or to establish population generalizability, we run the risk of assuming cultural differences when they are likely the result of a confounded variable. And having spent the greater part of my career working with Hispanic populations, I believe that these pseudo experiments and their erroneous conclusions have caused great harm and led to a general distrust of social science research and interventions within the Hispanic communities (Marín & Marín, 1991). My conclusion is that culture or ethnicity should not be used as an independent variable.

### **Definition of Constructs**

These confounded relationships are seen as threats to external validity in the traditional approach to validity of research design, which seeks to ascertain general laws that hold regardless of others variables. However, growing out of life-span developmental psychology, applied developmental science holds that confounded variables, or interactions, are the key to understanding human behavior (Hultsch & Hickey, 1978): Identification of the confounding variables and other dimensions of the context of research enhances the interpretation of the research findings.

This endeavor is difficult because the key constructs are often not adequately defined. For example, what is culture? Culture includes psychological aspects of shared norms, values, and beliefs along with material entities (such as paintings, music and dress) and the social structures of organizations and institutions (Fiske, Kitayama, Markus & Nisbett, 1998). To be of use to applied developmental science, culture must be defined as a process. In other words, specification of the

cultural context requires the explicit definition-- in research hypotheses – of the developmental processes that are assumed to be the mediating variables between culture and outcome (Busch-Rossnagel, 1992).

A similar approach should be taken to defining race or ethnicity if those concepts are used to define groups. In doing so, the likelihood of using panethnic labels such as Hispanic in place of Puerto Rican or Colombian, or Asian in place of Korean or Vietnamese, should diminish as national origin, acculturation, religion, immigration history, and other factors are identified to define ethnicity (APA, 2003; Fisher, Jackson, & Villarruel, 1997)). Likewise unraveling the confounds of ethnicity with educational attainment, income, employment status and so on can identify overgeneralizations that stigmatize individuals and ethnic groups (Sue, 1999).

### **Linguistic equivalence**

One process related to culture or ethnicity, namely linguistic competence, requires additional attention to promote culturally centered assessment. Most assessment is based on measures available in just English, but in today's diverse population, individuals participating in developmental assessments may not speak English or may not feel comfortable speaking it (Shin, 2003; U. S Census Bureau, 2004). The professional who wants to obtain information from a non-English speaking population is faced with creating measures that are linguistically equivalent.

This process usually starts with a simple translation, done by an individual with fluency in both languages. Translation is not an exact process, so it often includes having a number of bilingual individuals undertake the translation to achieve a consensus. Unfortunately, the result is likely to be phrases in the second language that are not true to the intent of the measure because of the lack of precision in the original version. The most precise words are often less commonly

used in everyday interactions, so their use in a psychological measure increases the reading level -- and the difficulty-- of the measure.

A second way of approaching translation is through the process of back translation, which is also known as double translation. For example, to create a Spanish version of an English measure, a bilingual person first translates the English version into Spanish, and this Spanish translation is then translated back into English by a second bilingual person. This completes the back- or double-translation. The back-translation (which is in English) is compared with the original English text. If the two versions are different, the Spanish version is altered to more closely approximate the original English. The altered Spanish version is then subjected to another back translation to English. Back translation through several iterations is usually seen as “best” practice to develop linguistically equivalent versions of measures. However, because only the Spanish version is modified, and the English version is not changed, back translation has limitations. When the original, English measure is standardized and cannot be modified without jeopardizing the psychometric information gathered on the standardized measure, then iterative back translation must suffice.

When both versions of the instruments are being developed simultaneously, a better option is available. This is the process of *decentering* (Werner & Campbell, 1970). On the surface, the process of decentering is the same as the iterative process of back translation. The difference is that when comparing versions, either the Spanish or the English version maybe modified to enhance the match between the two. Where discrepancies exist between the two versions, researchers can discuss the intent of the item, re-write the item for clarification in either language, and then translate and back-translate again. In this way, each round of translation informs the development process for both versions and often has the effect of clarifying the focus of the items.

Decentering is likely to affect the development of a measure because it clarifies the linguistic boundaries of the constructs and thus is likely to lead to culturally centered measures (APA, 2003; Rogler, 1999).

### **Valid Use**

In addition to design validity, construct validity, and linguistic equivalence, what else is necessary for applied developmental assessment to be valid? It is important to remember that an assessment (or a test) is not valid in and of itself; rather there is valid use of assessment (AERA, APA & NCME, 1999).

From this utilitarian, applied perspective, valid use first depends on the qualifications of the individual responsible for the assessment. Like traditional testing standards, applied developmental assessment assumes that the examiner is skilled, that s/he is able to establish rapport with the individual(s) being assessed, and that any assessments are administered, scored, and interpreted correctly.

In contrast to traditional approaches to testing, understanding of the bi-directionality of effects changes the role of this individual from that of the “expert” to that of a colleague. The participants in the assessment (and the guardians for children and youth) are equals in determining the valid use of the assessment. Most importantly, the participants are the ones who determine if assessment should occur and the utility of the information obtained. In other words, the participants help to establish the “referral questions” that define the assessment process. This perspective will be familiar to those involved in educational planning for children with disabilities, as Public Law 94-142, The Education for All Handicapped Children Act, helped to establish the importance of parents or guardians in the development of individualized educational plans for

their children. Applied developmental assessment expands this involvement further back into the assessment process.

In addition to a skilled examiner who respects the participants as contributors to the assessment process, the validity of assessment rests with an adequate examination of the contextual effects at all the possible levels. "Testing the limits" helps to identify context effects. By providing practice on specific skills or test items, testing-the-limits enlarges the context of assessment to observe what individuals can do under ideal, as well as under standardized, conditions (Kliegl, Smith & Baltes, 1989).

Other aspects of contexts may also be important. What is the effect of the context if early intervention screening is conducted in a medical clinic? What is the effect of having parent permission when the assessment is of the sexual behavior of adolescents? To identify contextual effects, context should be treated as one of the dependent variables to be examined by including multiple contexts during the assessment process.

Validity of use also requires an understanding of the lack of perfect prediction from the assessment result. While current functioning may be observed, future behavior is only inferred (Busch-Rossnagel, Hawryluk, & Pavone-Kennedy, 1983). And that inference occurs within the conceptual framework created by the assessment process. If the assessment professional has a thorough grounding in research methods, then the conceptual framework may be a series of competing hypotheses for which this assessment is looking for disconfirmation. Research methods teach us that we cannot confirm the null hypothesis, so we should phrase our assessment questions to identify and seek the information that will lead to us to reject our hypotheses. Unfortunately, many traditional assessment methods, especially projective tests (Anastasi & Urbina, 1997), approach the assessment process in the opposite way: they look for the

information consistent with the hypothesis rather than looking for discrepant evidence. Seeking disconfirming information often means looking for strengths—the positive behaviors—rather than just lengthening the list of behaviors the individual cannot perform.

### **Communication of results**

Finally, the utility of assessment rests with communication of the results, and here again best practice for applied developmental assessment differs from traditional approaches. In traditional practice the result of the assessment is a series of scores, usually standardized (e.g. an IQ). Because they are standardized (e.g., with a mean of 100 and a standard deviation of 15 for most IQs), they have meaning for the “experts.” Mental health professionals are socialized to write for other professionals, resulting in reports that are dense with jargon that may have little meaning and even less utility for the participants. Such reports are often focused on weaknesses, the areas of poor functioning that originally brought the participant in to the assessment process.

In contrast, the emphasis on the bi-directionality and collaboration in applied developmental assessments suggests that the participants should shape the communication of results. For some participants, a standard score may be understandable, but for many participants and guardians, the most understandable score is the percentile rank. Percentiles require fewer assumptions than standard scores and remove some of the mystique of scores in presenting only what a norm-referenced test can, which is the individual’s relative standing in a group (Salvia & Ysseldyke, 2001).

Quantitative information should be contextualized by identifying strengths as well as weaknesses as these may help identify ideas for improvement or enhancement. Narrative observations will also suggest strengths and ideas for promotion of positive behaviors. In this way

the communication of results from in depth assessment will logically lead to program planning and intervention.

### **Conclusion**

The need to communicate assessment results that include ideas for the promotion of development returns us to the heart of the process of applied developmental assessment. The goal of assessment should be to provide information that is useful to modify or optimize development. The life span perspective that underlies work in applied developmental science highlights that the potential for growth exists throughout the life course, but because of changes in the potential of plasticity across life, emphasizes the importance of early intervention to facilitate change most effectively (Baltes, Lindenberger & Staudinger, 1998). Applied developmental assessment will be most useful when coupled with early intervention at any age. Thus, my exploration of applied developmental assessment returns me to where I started out with Don in 1976, emphasizing the role of early intervention to optimize human development.

## References

American Psychological Association. (2003). Guidelines on multicultural education, training, research, practice and organizational change for psychologists. *American Psychologist*, 58, 377-402.

American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). (1999). Standards for educational and psychological testing. Washington, DC: American Educational Research Association.

Anastasi, A. & Urbina, S. 1997. Psychological testing, 7th Edition. Upper Saddle River, NJ: Prentice Hall.

Baltes, P., Lindenberger, U., & Staudinger, U. M. (1998). Life-span theory in developmental psychology. In W. Damon (Series Ed.) & R. M. Lerner (Volume Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development (5<sup>th</sup> Ed., pp. 1029-1144)*. New York: Wiley.

Busch-Rossnagel, N. A. (1981). Where is the handicap in disability?: The contextual impact of physical disability. In R. M. Lerner & N. A. Busch-Rossnagel (Eds.), *Individuals as producers of their development: A life-span perspective (pp. 281-312)*. New York: Academic Press.

Busch-Rossnagel, N. A. (1992). Commonalities between test validity and external validity in basic research on Hispanics. In K. F. Geisinger (Ed.), *Psychological Testing of Hispanics (pp.195-214)*. Washington, D. C.: American Psychological Association.

Busch-Rossnagel, N.A., Hawryluk, M.K., & Pavone-Kennedy, M. (1983). Issues and considerations in the early identification of exceptional preschool children. *New Jersey Journal of School Psychology, 2*, 2-8.

Fisher, C. B., Jackson, J. F., & Villarruel, F. (1997). The study of African American and Latin American children and youth. In W. Damon (Series Ed.) & R. M. Lerner (Volume Ed.), *Handbook of child psychology: Vol. 1. Theoretical models of human development (5<sup>th</sup> Ed., pp. 1145-1207)*. New York: Wiley.

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. T. Gilbert & S. T. Fiske (Eds.), *The handbook of social psychology (4th ed., Vol. 2, pp. 915-981)*. New York: McGraw Hill.

Francis, D., Schatschneider, C., & Carlson, C. (2000). Introduction to individual growth curve analysis. In D. Drotar (Ed.), *Handbook of research in pediatric and clinical child psychology (pp. 51-73)*. New York: Plenum Publishers.

Hultsch, D. F. & Hickey, T. (1978). External validity in the study of human development: Theoretical and methodological issues. *Human Development, 21*, 76-91.

Laosa, L. (1980). Maternal teaching strategies in Chicano and Anglo-American families: The influence of culture and education on maternal behavior. *Child Development, 51*, 759-765.

Kliegl, R. Smith, J. & Baltes, P. B., (1989). Testing-the-limits and the study of adult age differences in cognitive plasticity of a mnemonic skill. *Developmental Psychology, 25*, 247-256.

Lerner, R. M. & Busch-Rossnagel, N. A. (1981a). *Individuals as producers of their development: A life-span perspective*. New York: Academic Press.

Lerner, R. M. & Busch-Rossnagel, N. A. (1981b). Individuals as producers of their development: Conceptual and empirical bases. In Lerner, R. M. & Busch-Rossnagel, N. A. (Eds.),

*Individuals as products of their development: A Life-span perspective.* (pgs. 1-36). New York: Academic Press.

Marín, G. & Marín, B. V. (1991). *Research with Hispanic populations.* Thousand Oaks, CA: Sage.

National Commission on Excellence in Education. (1983). *A nation at risk.* Washington, D. C.: U. S. Department of Education.

Rogler, L. (1999). Methodological sources of cultural insensitivity in mental health research. *American Psychologist, 54,* 424-433.

Salvia, J. & Ysseldyke, J. E. (2001). *Assessment (8<sup>th</sup> ed).* Boston: Houghton Mifflin.

Shin, H. B. (2003). Census 2000 Brief: Language use and English speaking ability: 2000. Washington: U. S. Census Bureau. Retrieved April 4, 2004 from <http://www.census.gov/prod/2003pubs/c2kbr-29.pdf>.

Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist, 54,* 1070-1077.

U S Census Bureau (2004). Language Use. Retrieved April 4, 2004 from [http://www.census.gov/population/www/socdemo/lang\\_use.html](http://www.census.gov/population/www/socdemo/lang_use.html).

Werner, O., & Campbell, D. T. (1970). Translating, working through interpreters and the problem of decentering. In R. Cohen (Ed.), *A handbook of methods in cultural anthropology.* New York: American Museum of Natural History.

Wohlwill, J. F. (1973). *The study of behavioral development.* New York: Academic Press.