Early Education & Development
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t775653644

Formative Evaluation of the Early Development Instrument: Progress and Prospects

Daniel P. Keating [a]
[a] Center for Human Growth and Development, University of Michigan,

Online Publication Date: 11 October 2007
To cite this Article: Keating, Daniel P. (2007) 'Formative Evaluation of the Early Development Instrument: Progress and Prospects', Early Education & Development, 18:3, 561 - 570
To link to this article: DOI: 10.1080/10409280701610937
URL: http://dx.doi.org/10.1080/10409280701610937

EARLY EDUCATION AND DEVELOPMENT, *18*(3), 561–570

Formative Evaluation of the Early Development Instrument: Progress and Prospects

Daniel P. Keating
Center for Human Growth and Development
University of Michigan

This article is a commentary for the special issue on the Early Development Instrument (EDI), a community tool to assess children's school readiness and developmental outcomes at a group level. The EDI is administered by kindergarten teachers, who assess their kindergarten students on 5 developmental domains: physical health and well-being, social competence, emotional maturity, language and cognitive development, and communication skills and general knowledge. In this commentary, the author critically integrates research findings from projects that used the EDI to assess children's development at a community or population level, as presented in the research articles of this special issue. In addition, the EDI is situated in the school readiness literature, and forthcoming research directions and challenges that will largely determine the risk–benefit ratio of the EDI are discussed in regard to the following 3 topics: the advantages and limitations of using teacher judgment for the EDI, the establishment of the EDI's reliability and validity at a group (community, population) level, the EDI's effectiveness for intervention and program evaluation and measurement.

In an earlier special issue of this journal, Vernon-Feagans and Blair (2006) assembled a set of articles to probe the frontier of the field of school readiness and its assessment. For scholars who have not focused on this topic, to whom the concept may appear straightforward, the articles in that special issue offer an enlightening window into how complex and controversial the topic has become. Lacking this background, one might assume that the research field of school readiness is aimed at identifying the salient characteristics of children who are more versus less pre-

Correspondence should be addressed to Daniel P. Keating, University of Michigan, 300 North Ingalls, 10th Floor, Ann Arbor, MI 48109-0406. E-mail: keatingd@umich.edu

pared to benefit from standard school instruction, reliably and validly assessing those characteristics, and using the information gathered in this way to mitigate the limitations of children who are "less ready" by modifying aspects of developmental experiences prior to school entry and/or the early school experiences of such children so as to help them overcome their lack of readiness.

THE HISTORY OF SCHOOL READINESS

The historical perspective on readiness outlined by Vernon-Feagans and Blair (2006) and explored in greater depth by Graue (2006) suggests a number of ways in which these straightforward assumptions have proven to be, and continue to be, problematic. One key controversy inheres in the term itself. Do we mean the child's readiness to benefit from standard schooling? Do we mean the school's readiness to support all children, regardless of the diverse suites of abilities and limitations they possess? Do we mean society's readiness to support the development of children prior to school and during the transition to school such that children are well grounded in the essential skills for acquiring, during the balance of their school careers, the important proficiencies necessary for future social and economic participation in society?

From a developmental science perspective, it seems easy in principle (but difficult in practice) to answer all of these questions in the affirmative. We need to know the distribution of capabilities of children as they enter school; we need to know how to adapt curriculum and instruction (and the entire school experience) during the transition to school to accommodate that full distribution; and we need to know how best to provide the resources to accomplish both tasks, as a responsibility of the broader society—*if* we assume that the social benefits of population health and competence are sufficient to warrant adequate social investments to achieve those benefits (Keating, 1999).

The controversy arises because of differences in views about the causal chains that lead to variation in school transition success; about the locus of responsibility (or blame) for lack of success; and about the locus, nature, and size of social investments required to remedy the situation. If the primary diagnosis is that children are not ready for school, the focus of responsibility shifts to those whose task it is to get them ready, principally families and secondarily the communities that are implicitly expected to provide supports for early development. It is easy to see how this can quickly migrate toward blaming the victims, pointing to familial, community, and/or cultural factors that have done an inadequate job. In addition, it is easy to see how this can be seen as letting schools off the hook—"look at what we have to work with." If the primary diagnosis is that schools are not ready for all children, the focus of responsibility shifts to the schools to find ways (usually within existing budgets) to quickly bring all children, regardless of their limitations, to a foun-

dational developmental level that enables them to acquire academic proficiency via typical school instruction. In this view, the schools are to blame, which analogously lets families, communities, and the broader society off the hook for failing to ensure the developmental opportunities needed to prepare children to benefit optimally from formal education. If the primary diagnosis is that society has failed to invest adequately in the continuum of care from infancy through early childhood and in schools' ability to accommodate the full range of diversity, then the responsibility seems to be fully diffused, and the particular targets for wise investments remain unspecified.

Confronted with this circle of blame, what is the way forward? Graue (2006) argued that all actors should address their respective components: "Readiness is an ethical responsibility we have to children that encompasses coordinated systems of early care and education and receptive schools that are developmental, inclusive, and accountable to all" (p. 43). Accepting this responsibility suggests that advances in research that can support evidence-based policy and practice are essential in each arena. In the previous special issue (Vernon-Feagans & Blair, 2006), a range of relevant research was reported, including core considerations for the improvement of readiness assessment in general (Snow, 2006), of socioemotional development (Denham, 2006; Ladd, Herald, & Kochel, 2006), of preliteracy skills (Lonigan, 2006), and of the instructional and social components of early instruction (Dickinson, 2006). Mashburn and Pianta (2006) argued that the interactions and transactions amongst all of these factors are central to improving school readiness, and these authors emphasized the centrality of understanding and improving the social relationships that are the primary mediators of readiness in order to provide "guidance for designing interventions to improve school readiness through strengthening relationships between parents and children, parents and teachers, and teachers and children" (p. 151).

Given the scope of the topics that are brought under the umbrella of school readiness by attending to all of these factors, there exists a risk that the complexity could become overwhelming. Where exactly should we start? A sensible first step is to refine the tools available for assessment of the current situation. Snow (2006) provided an overview of key issues and argued that it is central to keep the purpose of assessment firmly in mind. If the purpose is to provide a diagnosis of an individual child for the purpose of placement or instructional decisions, for example, the measurement burden is different from the purpose of providing a classroom- or school-level overall assessment of readiness status.

THE EARLY DEVELOPMENT INSTRUMENT (EDI)

These issues provide the background for the current special issue that focuses on the use of a relatively new comprehensive assessment of children's readiness, the

EDI (Guhn, Janus, & Hertzman, this issue; Janus & Offord, 2007). Several characteristics of this instrument, taken together, make it an innovative and potentially valuable measurement tool in this field. First, it is intended specifically to be used as a community-level assessment to aid in policy and planning, and it is not intended to be used at the level of individual assessment. This eases some measurement burdens with respect to diagnostic and/or screening accuracy, for example, but entails a new set of measurement assumptions that are likely to prove equally challenging in their own right. Second, it makes use of teacher judgments (typically toward the end of the kindergarten year), which makes it a cost-effective methodology for acquiring aggregated information that can be used at a population level. Third, it includes five major domains that have been found in a range of studies to be important for subsequent success in school, although there is some evidence that multidimensional models may overidentify children at risk (Crooks & Peters, 2005), that is, generate an unacceptably high rate of false positives. Given that the EDI is not intended for individual screening, this may be less of a concern, although the problem of policy inferences based on overidentification of risk may be nontrivial.

The series of articles based on research applications of the EDI included in this special issue begins to address the complex questions of science, policy, and practice that arise in using the EDI. A number of articles take up standard measurement criteria of reliability and validity, for the most part successfully. Another set of articles deals with some of the unique characteristics of the EDI, particularly patterns of neighborhood variability that can be conceptualized as a different version of validity, namely its validity as a population measure. These findings provide initial support for this intended use of the EDI but also raise a number of challenging questions that will need to be addressed in future research. Finally, the issue of how effective the EDI is for its intended use as a community-level planning and assessment tool is joined in several articles, but the evidence on this critical aspect is mostly provisional at this time. Some issues that will need to be addressed to realize this final goal are identified below.

MAXIMIZING THE EDI BENEFIT–RISK RATIO

The potential benefits of the EDI are described in various contributions to this special issue. They fall into three broad categories. First, because the methodology for deriving information about children's readiness for school is obtained through teacher reports on individual children, the cost relative to standardized individual assessments is substantially lower. This makes feasible the collection of data at aggregated population levels, affording a different level of analysis. It can also be done on a large scale on a routine basis, affording the study of change at various levels of aggregation, from neighborhood to community to municipality or larger

jurisdictions. In addition, it can be linked to other population-level databases to provide opportunities for analyses of multiple contributions or barriers to school readiness. Second, the EDI collects data across five broad domains that prior research has linked to subsequent school success, which offers not only a broad portrait of school readiness but also an opportunity to look at discrepancies among domains of readiness at an aggregated level. Third, the outputs from EDI reporting can be organized in such a way that they are potentially useful for local and community planning to address targeted areas of concern, a goal that has been identified for some time but has proven difficult to realize (Keating & Mustard, 1996).

The potential benefits of the EDI do not come without risks, of course. The first step in guarding against those risks is to identify them, and the second is to carry out well-designed research studies that minimize the potential risks. Not all risks, or benefits, can be identified at the launch of a research program, but some key ones can be anticipated. The articles in this special issue provide an opportunity to evaluate the evidence thus far on the EDI and to point toward future, difficult questions that need to be answered for this tool's potential to be realized.

The Methodology of Teacher Judgment

There is the possibility that the limitations of the methodology, teacher report, may become lost over time. This risk of reification, treating the measurement outcomes as if they are real instantiations of the construct (readiness in this case), is not unique to the EDI, of course. It has been a persistent problem for theory, research, and practice throughout the history of measurement (Keating, 1990). But there are some special characteristics of the EDI that warrant close attention.

As teachers render their judgments on individual children in their classrooms, how strongly their social assumptions about the population of children they are serving enter into those judgments may prove hard to assess. This is one of the major reasons that "objective" testing, either in standardized group tests or clinical assessment, was introduced in the first place, precisely as an attempt to create safeguards against such bias. This is not to argue that such bias is either motivated or conscious. There is by now a large research literature to support the view that such biases can enter into social comparative judgments outside of the judger's awareness (Chambers & Windschitl, 2004). In addition, teacher judgments have been found to be reliable even at the individual level, although in one validated approach a more detailed (and expensive) assessment protocol was employed (Meisels, Bickel, Nicholson, Xue, & Atkins-Burnett, 2001).

One may argue that the risk of biased judgments pertains primarily in the case of decision making about individual children, for which the EDI is not intended to be used. But of course the validity of the EDI at a population level could be similarly compromised. Because understanding the social patterning of school readiness is a primary goal of the EDI, it is possible that spurious correlations between

the EDI and neighborhood or community characteristics could arise from the same processes.

Reliability and Validity: Individual and Group

One of the main challenges for the EDI is how best to establish that it meets the standard criteria of measurement quality, reliability, and validity. Although this is true for any new psychometric assessment tool, there are additional challenges for the EDI. Even though the measurement burdens are in some respects less onerous for the EDI in that it has not been designed to serve as an individual readiness indicator or clinical diagnostic tool, there are some novel challenges. What does reliability and validity mean for a *group* readiness indicator? If the EDI is to be used as intended (i.e., as a means for assessing aggregate differences among neighborhoods, communities, and so on, and especially as a measure of naturally occurring or induced change across time), then its suitability for that role requires close examination. The criteria for reliability and validity in this sense are not nearly as well established as for classically understood reliability and validity and will likely necessitate innovative techniques that remain to be fully developed.

How has the EDI fared thus far in guarding against these measurement challenges? One design feature was to make the judgments as specific as possible by asking about particular behaviors rather than about general impressions. One early test of this is whether the standard reliability assessments yield the predicted measurement model. The analyses carried out during the instrument development phase of the EDI yielded positive results on confirmatory factor analyses as well as substantial internal consistency reliabilities, as expected: physical health and well-being, 0.84; social competence, 0.96; emotional maturity, 0.92; language and cognitive development, 0.93; and communications skills and general knowledge, 0.95 (Janus & Offord, 2007, p. 10). Further analyses suggested that the observed variation was primarily among children rather than between classrooms or teachers, again as expected. Taken together, the initial evidence suggests that the EDI functions reliably at the individual level and is not based on global teacher impressions.

It is less clear how measurement reliability at the group level is to be established. It is not immediately obvious what the analogue of internal consistency at the group level would consist of. One might argue that the scaling up from individual-level internal consistency is sufficient, but this is not self-evident. Not only is the methodology of teacher judgment implicated in any individual internal consistency assessment—having already rated a child as lacking skills in a specific domain, it is hard to imagine setting this judgment aside sufficiently to avoid its halo effect on one's judgment of a different within-domain skill—the idea that internal consistency automatically transfers to the instrument when it is used at the group level would seem to entail a connected argument, rather than be based on assump-

tion. Similar cautions would seem to be appropriate if one were to turn toward recent developments in item response theory as an approach, namely, whether and how it could be adapted to assessing aggregate reliability of measurement.

The force of this somewhat arcane revisiting of classic reliability questions (Stanley, 1971) is perhaps more compelling when we consider a more transparent variety of reliability, test–retest reliability. Here the contrast between individual-level reliability and group-level reliability is somewhat clearer. Test–retest reliability at the individual level is intended to indicate that the same (or a parallel) measure administered after a reasonably short interval will yield the same result, suggesting that the yardstick is not overly flexible. This is easily understood, and in the case of the EDI we should expect fairly high levels of test–retest reliability, given that teachers are being asked again about the judgments they just recently made. In fact, this result is reported by Janus and Offord (2007, p. 16).

At the group level, however, we confront a somewhat different question. What does it mean to have a reliable yardstick of aggregate-level readiness? At one level, we could mean merely the individual test–retest reliabilities, aggregated by group and then contrasted across groups. It seems almost certain that this would be strongly positive, given that group means are always more stable than individual scores. But for the EDI to work as intended, it seems clear that there are additional criteria. What is needed is a yardstick that assesses stability of background measurement at the group level, such that patterns of change, either naturally occurring or as the result of planned interventions at the community level, can be adequately assessed.

Another protection against measurement risk is to examine the concurrent validity of the EDI against other indicators of readiness and its predictive validity with respect to subsequent school achievement. Both questions were examined in several of the article in this special issue. Generally speaking, the EDI correlates well with other ratings of readiness (Brinkman et al., this issue) and was also found to be as predictive of achievement at the end of first grade as were direct assessments of children (Forget-Dubois et al., this issue).

The concurrent validity of the EDI to direct performance assessments (as opposed to parent or teacher ratings) for individuals was not reported in the assembled articles, but Lesaux, Vukovic, Hertzman, and Siegel (this issue) reported such correlations for school-aggregated EDI and direct literacy assessments. This is especially pertinent information, in that it deals with the intended use of the EDI as an aggregate measure. Each of 23 schools received a mean literacy score for its kindergarten children ($N = 478$), for letter–word identification and for phonological awareness. For each of the five EDI scales, the percentage of children deemed to be at risk (a rating in the bottom 10% of the full population in the respective domain) in each school was also calculated. The pattern of results at the aggregated level showed some substantial correlations but also revealed some puzzling findings. Of the four correlations that would seem to be most pertinent, EDI Language

and Communication with letter–word identification and phonological awareness, only one was significant ($r = –.51$, $p < .05$; Lesaux et al., this issue, Table 2). The largest absolute correlation was between letter–word identification and EDI Physical Health and Well-Being. This puzzling finding is somewhat resolved in noting the types of items that are included on that EDI scale ("How often has the child arrived to school too tired to do school work?" and "How would you rate this child's level of energy throughout the school day?"; Janus & Offord, 2007, p. 6, Table 1), which may suggest that teachers are noting some substantial general problems with the child's ability to even participate in the school experience. These may reflect physical health problems or pervasive family stress or family functioning issues—a classic psychometric "jingle-jangle" fallacy.

It is exactly this type of finding that will not only pose challenges for interpretation of school-level readiness but will also make it a nontrivial issue of designing prevention or intervention measures. The EDI scales focusing on the performance domain are not picking up the reading difficulties (at least at the aggregated level), but the EDI is picking up some apparently generalized problems with children's ability to engage with school. Should scarce resources be directed primarily toward the encouragement of parental reading practices or home-based cognitive stimulation, toward more aggressive intervention in preschool and kindergarten classrooms to provide a remedial foundation, or toward reducing family stress and increasing family functioning such that children are able to benefit from what is going on during the school day? All may be desirable, of course, but the promise of the EDI is to aid in community planning, which seems to require more precise diagnostics than "do everything." Future research will need to proceed along a number of avenues to unravel these and similar conundrums.

It is clear that the EDI does relate to other background variables, including socioeconomic status of the families, in ways that we would expect, and that reinforces its convergent validity (Janus & Duku, this issue). It is not yet clear whether the EDI is tapping variance in readiness that is the same as that measured in other ways, or whether it is perhaps tapping variance that has not been as apparent in prior research. Only a substantial line of research to explore the various empirical possibilities is likely to answer that important question.

The validity of the EDI at the aggregate level is considered further in several articles on the influence of neighborhood variables after controlling for individual-level socioeconomic contributions (Lapointe, Ford, & Zumbo, this issue; Lesaux et al., this issue). These patterns need to be considered in light of the methodology of teacher judgment, as noted previously. There is evidence for discriminant validity in that different EDI domains show differential patterns despite the fact that it functions similarly across groups in general (Guhn, Gadermann, & Zumbo, this issue), although in largely expected ways, such as teacher judgments about language in schools that serve large numbers of English-language learners. Nevertheless, these aggregate patterns of convergent and discriminant validity es-

tablish an important baseline that warrants further research of a promising new methodology.

Effectiveness of Community Interventions

The reason for dealing at length with issues of *group-level* reliability and validity is that they are critical to the promise of the EDI as a tool in the assessment of effective community-level interventions to improve readiness. Three key questions arise here. First, what counts as "real" change? That is the core measurement issue. As the EDI becomes more embedded as a community-level planning tool, the issue of what it can reveal about the impact of community change will come to the forefront (Kershaw, Forer, Irwin, Hertzman, & Lapointe, this issue; Sayers et al., this issue). Here, the promise of administrative linkage using a full population health model is hopeful, in that it can potentially provide a unique window into broad issues of aggregate reliability and validity (Kershaw et al., this issue).

Second, and likely to prove even more vexing: What is an adequate design for assessing the active components of an intervention or prevention effort, given the great difficulty of community-based randomized control trials? This is a difficulty nearing impossibility when the challenges of enforcing provisions against "contamination of best practices" is taken into consideration on the timelines required for community-level interventions. Some interveners would argue that randomized control trials are the wrong model for examining community-level interventions for these and other reasons, but the burden will rest heavily on these experts to design evaluations that are capable of rigorously assessing such interventions, sufficient to merit being regarded as evidence based. There may be some useful borrowing from new developments in clinical intervention science that focus on identifying the active components of interventions, rather than a yes/no answer (Collins, Murphy, Nair, & Strecher, 2005). But these and other evaluation designs will depend on the EDI providing reliable and valid indexes of domain-specific change, as already noted.

One final question arises in terms of how the EDI will function if it is drawn into the now pervasive accountability movement—so pervasive that some might regard the question as when it will be drawn in, rather than if. For example, how is "gaming the system" to be prevented, either as a halo effect (intervention participants want it to succeed and may increase EDI ratings, consciously or not) or, even more forcefully, if EDI were to become a "teacher accountability" criterion. Perhaps these future possible pitfalls can be avoided, but to do so will require considerable forethought.

It is fair to say, then, that significant work remains for the EDI to realize its potential, and this work needs to proceed in parallel to deal with measurement issues, validity issues, implementation issues, and evaluation issues. What the current collection of articles strongly indicates, however, is that there are already substantial,

even impressive, programs of research to address these questions. This is a desirable development, because the importance of school readiness is such that having such an instrument in the toolkit will clearly improve the chances of acting successfully on that agenda.

REFERENCES

Chambers, J., & Windschitl, P. (2004). Biases in social comparative judgments: The role of nonmotivated factors in above-average and comparative-optimism effects. *Psychological Bulletin, 130,* 813–838.

Collins, L., Murphy, S., Nair, V., & Strecher, V. (2005). A strategy for optimizing and evaluating behavioral interventions. *Annals of Behavioral Medicine, 30*(1), 65–73.

Crooks, C., & Peters, R. (2005). Predicting academic difficulties: Does a complex multidimensional model outperform a unidimensional teacher rating scale? *Canadian Journal of Behavioural Science, 37*(3), 170–180.

Denham, S. A. (2006). Social–emotional competence as support for school readiness: What is it and how do we assess it? *Early Education and Development, 17,* 57–89.

Dickinson, D. K. (2006). Toward a toolkit approach to describing classroom quality. *Early Education and Development, 17,* 177–202.

Janus, M., & Offord, D. (2007). A teacher-completed measure of children's readiness to learn at school entry. *Canadian Journal of Behavioural Science, 39*(1), 1–22.

Graue E. (2006). The answer is readiness–Now what is the question? *Early Education and Development, 17,* 43–56.

Keating, D. P. (1990). Charting pathways to the development of expertise. *Educational Psychologist, 25,* 243–267.

Keating, D. P. (1999). Developmental health as the wealth of nations. In D. P. Keating & C. Hertzman (Eds.), *Developmental health and the wealth of nations: Social, biological, and educational dynamics* (pp. 337–347). New York: Guilford Press.

Keating, D. P., & Mustard, J. F. (1996). The National Longitudinal Survey of Children and Youth: An essential element for building a learning society in Canada. In *Growing up in Canada: National Longitudinal Survey of Children and Youth* (pp. 7–13). Ottawa, Ontario, Canada: Human Resources Development Canada and Statistics Canada.

Ladd, G. W., Herald, S. L., & Kochel, K. P. (2006). School readiness: Are there social prerequisites? *Early Education and Development, 17,* 115–150.

Lonigan, C. J. (2006). Development, assessment, and promotion of preliteracy skills. *Early Education and Development, 17,* 91–114.

Mashburn, A. J. & Pianta, R. C. (2006). Social relationships and school readiness. *Early Education and Development, 17,* 151–176.

Meisels, S. J., Bickel, D. D., Nicholson, J., Xue, Y., & Atkins-Burnett, S. (2001). Trusting teachers' judgments: A validity study of a curriculum-embedded performance assessment in kindergarten to grade 3. *American Educational Research Journal, 38,* 73–95.

Show, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early Education and Development, 17,* 7–41.

Stanley, J. C. (1971). Reliability. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Vernon-Feagans, L., & Blair, C. (2006). Measurement of school readiness. *Early Education and Development, 17*(1), 1–5.