



Examining the psychometric properties of the Infant–Toddler Environment Rating Scale–Revised Edition in a high-stakes context

Rossana Bisceglia^{a,*}, Michal Perlman^a, Diana Schaack^b, Jennifer Jenkins^a

^a Department of Human Development and Applied Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1V6

^b Erikson Institute, Loyola University Chicago, Chicago, IL 60611, USA

ARTICLE INFO

Article history:

Received 6 October 2007

Received in revised form 9 February 2009

Accepted 11 February 2009

Keywords:

ITERS

ITERS-R

Psychometric Properties

Factor Analysis

High-stakes

Staff-to-child ratio

Child care quality

One-classroom score

Multilevel modeling

Intra-class Correlation

ABSTRACT

The psychometric properties of the Infant–Toddler Environment Rating Scale–Revised Edition (ITERS-R) were examined using 153 classrooms from child-care centers where resources were tied to center performance. An exploratory factor analysis revealed that the scale measures one global aspect of quality. To decrease redundancy, subsets of items were selected randomly and by experts who rated items according to ease of administration and importance to quality. The shorter subsets demonstrated good discriminant validity, adequate to good psychometric properties, and high associations to the full ITERS-R score. They also demonstrated similar associations to staff education and staff-to-child ratio, as the full instrument. The best assessment of quality was demonstrated by the shortened subset that included items that assess both structural and process features of quality. Multilevel-analyses indicated that classrooms from the same providers score more similarly on ITERS-R than classrooms from other providers. The implications for using the ITERS-R in high-stakes contexts are discussed.

© 2009 Elsevier Inc. All rights reserved.

More and more infants and toddlers are being cared for outside of their homes. In 2005, approximately 20% of American children under the age of two attended regulated early child-care programs (*America's Children in Brief: Key National Indicators of Well-Being, 2006*). Research on the effects of child-care on children's development has shown that the quality of care matters to children's well being. Higher quality care is associated with positive developmental outcomes, both in the short-term (Burchinal, Roberts, Nabors, & Bryant, 1996; Dunn, 1993; Peisner-Feinberg & Burchinal, 1997) and long-term (Broberg, Wessels, Lamb, & Hwang, 1997; Campbell & Ramey, 1995; Currie & Thomas, 1995; Howes, 1988; Peisner-Feinberg et al., 2001), encompassing a wide range of developmental domains including cognitive functioning, language development, social competence, and emotional adjustment (e.g. Howes, 1988; National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network, 2000; Peisner-Feinberg et al., 2001). Despite the consensus that quality matters, the definition and measurement of quality remain elusive.

In general it is agreed that quality includes both “structural features” such as staff-to-child ratio and group size, and “process features” which involve the interactions and experiences that children have in the child-care center (Goelman et al., 2006). Measurement approaches that attempt to assess overall or global quality include indicators of both structural and process features. The *Infant–Toddler Environment Rating Scale* (ITERS; Harms, Cryer, & Clifford, 1990) is a widely used instrument in research on child care quality (Blau, 2000; Campbell & Milbourne, 2005; Gevers Deynoot-Schaub & Riksen-

* Corresponding author.

E-mail address: rbisceglia@oise.utoronto.ca (R. Bisceglia).

Walraven, 2006; Goelman et al., 2006; Herrera, Mathiesen, Merino, & Recart, 2005; Scarr, Eisenberg, & Deater-Deckard, 1994; van Ijzendoorn, Tavecchio, Stams, Verhoeven, & Reiling, 1998). The instrument assesses overall or global quality of the classroom environment for children up to 30 months of age. This study presents the psychometric properties of the revised edition of the ITERS (ITERS-R; Harms, Cryer, & Clifford, 2003).

Research on the measurement properties of instruments such as the ITERS-R has become increasingly important given the recent funding initiatives in the United States and to a lesser extent in Canada to motivate improvements in the child-care experience (Zellman & Perlman, 2008; Zellman, Perlman, Le, & Setodji, 2008). These initiatives tie the funding that a child-care center receives to the facility's scores on quality measures. Hence, scores on instruments such as the ITERS-R may increase or decrease a facility's funding. Other initiatives include tying teacher salaries and bonuses to classroom scores on the evaluation instrument, and making scores public so that parents can use them in selecting care for their children.

To our knowledge, no research has examined the validity of using the ITERS-R to make high-stakes decisions. Yet, according to the *Standards for Educational and Psychological Testing* (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 1999) instruments should be administered in contexts and for purposes for which they have been validated. A fundamental purpose of test validation is to examine the positive and negative consequences of test administration, and to determine that specific benefits will likely be realized by the administration of the test.

It is unknown whether the abovementioned practices of tying funding and salaries to ITERS-R scores improve the child-care experience. Conversely it is possible that the contingencies may motivate some child care providers to introduce facile changes to the center, such as modification to the space and furnishing and the arrangement of the classroom, in order to increase their scores on the instrument. If the scale does not measure distinct aspects of quality, such changes may increase the facility's overall score while possibly over-estimating the quality provided by the center. Indeed, research on high-stakes testing in school settings shows that as teachers learn the assessment systems, they modify their classroom curriculum and arrangement in keeping with the content of the evaluation tool (Corbett & Wilson, 1991; Shepard & Dougherty, 1991). Although this may result in improvements in the teaching of certain constructs and skills, concepts that may be more difficult to measure or that are not consistent with the operational definition of quality incorporated in the instrument may not receive adequate attention. We infer from this research that child care providers may respond similarly as elementary and secondary teachers to such high-stakes contexts, rendering the psychometric properties of the instrument particularly important.

Further, the lengthy administration of the ITERS-R, which spans from 4 to 5 h per classroom, may limit the use of the instrument by large-scale agencies, for large research projects and community uses, which require time-efficient instruments. The lengthy administration also raises questions with respect to the allocation of funds and the cost associated with the assessments. Perhaps the funds used in assessment would be better spent on quality improvements in centers that have limited resources. Identifying strategies to reduce evaluation cost can increase the use of the instrument in both research and applied settings.

One strategy to reduce assessment burden could be to administer the instrument to fewer classrooms per center. Research on the psychometric properties of the Early Childhood Environment Rating Scale-Revised (ECERS-R; Harms, Clifford, & Cryer, 1998) which is a sister measure to the original ITERS used in classrooms serving preschool-age children, indicates that classrooms from the same provider tend to score similarly to one another on the instrument compared to classrooms from different providers (see Perlman, Zellman, & Le, 2004). This suggests that the assessment of a few classrooms per center may be as reliable as assessing all classrooms from the same provider. In this paper, we evaluate the validity of this strategy by examining the extent to which the variance in classrooms' ITERS-R scores is attributable to provider membership. To our knowledge no research has empirically examined this question.

An alternate evaluation strategy could be to administer a shorter version of the ITERS-R. However, this strategy would only be valid if the instrument measures one global factor and if the items are highly correlated with one another which would suggest that a shorter version of the scale may capture the construct just as well as the longer version.

In the present study, we report results from an exploratory factor analysis based on the ITERS-R scores from 153 classrooms where funding allocation is tied to the scale's scores. We assess the validity of several cost-saving strategies for assessment, in a similar way as Scarr and colleagues did for the ITERS. Scarr et al. (1994) examined the psychometric properties of three subsets of 12 randomly selected items of the ITERS. The subsets demonstrated psychometric properties similar to the full scale. The authors concluded that the single dominant factor assessed by the full ITERS could be measured with good reliability and validity by using a shorter version. Further, Scarr et al. (1994) suggested that *less* than 12 items may be sufficient in capturing overall quality, and that researchers could select items that are consistent with their own theoretical and methodological purpose. Similar findings were reported for ECERS-R (Perlman et al., 2004). Based on those findings several research efforts have used a shortened version of the ECERS-R (Karoly, Ghosh-Dastidar, Zellman, Perlman, & Fernyhough, 2008). In this paper, we examine the validity of the suggestions made by Scarr et al. (1994) by comparing the psychometric properties of shorter subsets to the full ITERS-R instrument.

We also examined the extent to which the subsets, classified classrooms to the same quality categories as the full instrument. To further test the validity of the subsets, we examined the associations among the score of the full instrument and each subset, and the extent to which the full instrument and the subsets showed similar associations to regulatable indicators predictive of quality. If the subsets demonstrate comparable psychometric properties to the full instrument, researchers and policymakers may choose to administer shorter subsets to reduce time and cost associated with evaluation, while maintaining reliability. Since becoming a reliable rater on the full instrument requires time and frequent

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات