



Evaluating construct validity and internal consistency of early childhood individualized family service plans



Lee Ann Jung^a, Kelly D. Bradley^{a,*}, Shannon O. Sampson^b, Robin A. McWilliam^c, Michael D. Toland^a

^a University of Kentucky, United States

^b Independent Researcher, United States

^c Siskin Childrens Institute, United States

ARTICLE INFO

Article history:

Received 25 March 2014

Received in revised form 15 January 2015

Accepted 2 March 2015

Available online 28 March 2015

Keywords:

Early childhood

IFSP

Measurement

Survey research

ABSTRACT

This study presents evidence regarding the construct validity and internal consistency of the IFSP Rating Scale (McWilliam & Jung, 2001), which was designed to rate individualized family service plans (IFSPs) on 12 indicators of family centered practice. Here, the Rasch measurement model is employed to investigate the scale's functioning and fit for both person and item diagnostics of 120 IFSPs that were previously analyzed with a classical test theory approach. Analyses demonstrated scores on the IFSP Rating Scale fit the model well, though additional items could improve the scale's reliability. Implications for applying the Rasch model to improve special education research and practice are discussed.

© 2015 Elsevier Ltd. All rights reserved.

Introduction

In 1986, the Individuals with Disabilities Education Act (IDEA) (P.L. 99-457) was amended to include a section regarding Early Intervention (Part C), which includes the provision of developmental and related services to infants and toddlers, birth to 3 years, who have developmental delays or disabilities. Even though states are not required to participate in Part C, all states have elected to do so. By choosing to participate, and in turn, receiving federal funds to support early intervention programs, states must adhere to the regulations set forth by Part C in IDEA. The regulations provide a number of rights for families receiving federally funded early intervention. Each family is guaranteed a team-developed, legally binding individualized family service plan (IFSP) that outlines the specific services and supports that the family receives. The IFSP includes information about the child's present level of development, family priorities and concerns, outcomes for the child and family, and strategies the intervention team will use. The purpose of this study is to evaluate the measurement instrument constructed to rate the quality of the written IFSP document.

Though federal legislation requires states to include the above components, there has been little effort to systematically measure

the quality of these documents at the federal or even state level (Jung, 2010). Currently, states may develop their own forms or even allow districts or programs within their states to maintain unique forms for this purpose. A validated instrument that could guide systematic review and feedback on IFSP quality could prove useful in improving the plans and services provided to young children with disabilities or developmental delays. A standard rating scale could also be useful to service coordinators who wish to engage in self-evaluation by objectively examining their own IFSPs. To this end, McWilliam and Jung (2001) constructed the individualized family service plan *IFSP Rating Scale* in an effort to create a more standardized measure of the quality of these plans, which are written for infants and toddlers with disabilities.

Studies have been conducted utilizing traditional statistical approaches, classical test theory (CTT) such as principal components analysis, to examine the structure of items on the IFSP rating scale (Jung & McWilliam, 2005). Even so, additional understanding of response patterns and an examination of theoretical structure through an empirical measurement lens are needed to further the development of the scale. Because IFSPs play a fundamental role in planning early childhood special education services, it is critical to establish evidence on the construct validity and internal consistency of scores produced by the IFSP rating scale. Given the previous findings of studies analyzing the IFSP scale taking a CTT approach, the scale was assumed to serve as a unidimensional measure. The Rasch measurement model is the ideal tool to

* Corresponding author. Tel.: +1 859 257 4923; fax: +1 859 257 4243.

E-mail address: kdbrad2@uky.edu (K.D. Bradley).

confirm, or challenge this assumption, evaluate the structure of the scale, and offer suggestions for improvement and support for implementation of the measure.

Although estimates of inter-rater agreement, internal consistency, and component structure in previous studies through CTT support the notion that quality of IFSPs can be reliably measured using the IFSP Rating Scale (Jung, 2010; Jung & Baird, 2003; Jung & McWilliam, 2005), application of the Rasch measurement model to the scale can guide refinement of the rating scale and provide additional evidence of construct validity. The Rasch model, introduced by Georg Rasch (1960); Georg Rasch (1980), yields a comprehensive picture of the construct being measured and the concurrent responses, as the amount of a given latent trait in a person and the amount of that same latent trait reflected in various items can be estimated independently yet still compared explicitly. Simply stated, it provides diagnostic information on both respondents and items, allowing for a thorough review of the quality of the instrument. The Rasch model is a one-parameter logistic model within item response theory (IRT) family. The application of the model allows observations of respondents and items to be connected in a way that indicates the occurrence of a certain response as a probability rather than certainty and maintains order in that the probability of providing a certain response defines an order of respondents and items (Wright & Masters, 1982). Given that the Rasch model follows mathematically from the requirement of invariance of comparisons among persons and items, a Rasch analysis is appropriate when the total score on a questionnaire is used to make inferences about the level of a latent trait inherent. In this study, the latent trait would be the level of quality of the written IFSP. While a traditional statistical approach also uses the total score to characterize each respondent, it asserts the total score as the relevant statistic with little to no consideration of anomalies in the items and/or the respondents. The Rasch model accounts for these anomalies and provides a more informative score (Andrich & Luo, 2003). In this confirmatory study, the investigators apply the partial credit Rasch measurement model to the *IFSP Rating Scale* (McWilliam & Jung, 2001) using the 120 IFSPs previously analyzed by Jung and Baird (2003) and Jung and McWilliam (2005) to evaluate the construct validity and internal consistency of instrument.

Method

Response frame

The IFSPs used in this study are from one state. Even though early intervention policy varies from state to state, a study of a single state has great value, especially when considering an exploration and validation of a measurement tool. Using this response frame, a ‘ruler’ is constructed that can then be applied to similar interventions in other states. The utility of application is supported through consistent findings with previous research using the same scale in a different state (Jung, 2010). While the response frame is not a random sample of all IFSPs, it is a representative reflection of data collected using the instrument. More so, a highlighted strength of the Rasch model is that results are not constrained to be sample dependent, as is the case with classical test theory. Rasch analysis is independent of the distribution of the sample so long as the data collected is of a reasonably similar nature (Rasch, 1980). Thus, the response frame presented serves as a stable calibration sample.

Participants were service coordinators in one U.S. southeastern state’s early intervention system. As a part of a self-study conducted by the state early intervention system, IFSPs were requested by state administrators from the state’s 300 early intervention programs and district early intervention coordinators.

Following this request, 135 service coordinators voluntarily provided one IFSP and a non-identifying demographic survey for inclusion in the study. For calibration purposes, a second rater scored 15 IFSPs, and inter-rater agreement was high and reliability of the ratings was confirmed. For this study, only the ratings of the principal rater were included in the analysis, resulting in a response frame of 120 IFSPs. Because service coordinators were instructed to remove all identifying information from these IFSPs to protect child, family, and provider confidentiality, it was not possible to determine which service coordinators responded.

Overall descriptive statistics were produced using the demographic survey responses. Of those responding, service coordinators’ experience ranged from one month to over 12 years with a mean of three years. More than half of service coordinators had less than two years of experience. Just over half (53%) reported education degrees, 31% had education-related degrees and 16% had degrees unrelated to education. Most service coordinators reported having bachelor’s (61%) or master’s (32%) degrees; 7% reported having associate’s degrees. Of the IFSPs submitted, 71% were written by service coordinators who had recently attended a two and one-half day workshop on family-centered practices in natural environments.

Instrumentation

The *IFSP Rating Scale* (McWilliam & Jung, 2001) is an instrument that was developed to assess the quality of IFSPs. The instrument is an adaptation of McWilliam’s (1993) earlier version and contains items selected based on studies of families’ reactions to intervention plans (e.g., Able-Boone, Sanridall, Loughry, & Frederick, 1990; Summers et al., 1990), reflective writings about family-centered intervention plans (e.g., Bailey et al., 1986; Boone, McBride, Swann, Moore, & Drew, 1998; Johnson, McGonigel, & Kaufmann, 1989), curricula for developing family-centered intervention plans (Giangreco, Cloninger, & Iverson, 1993; McWilliam, 1992), and a review of the literature on family-centered practices in natural environments. Some of the items represent the more basic indicators of quality (selecting outcomes that match families’ priorities), and others are indicative of the higher level qualities (strategies embedded in routines). The idea is to apply this instrument as an evaluation tool. Thus, here a calibration response frame is used to investigate the fit and function of the tool itself, the *IFSP Rating Scale*.

An evaluator assigns a rating to each item on the *IFSP Rating Scale* according to the quality represented in a given IFSP. Each item is rated on a 5-point scale from 1 (*with descriptions of characteristics least consistent with family-centered practice*) to 5 (*with descriptions of characteristics most consistent with family-centered practice*). The rating manual, which accompanies the instrument, includes directions and examples for determining ratings of 1, 3, or 5. Ratings of 2 and 4 are used when the rater has difficulty determining which rating to assign. For example, if a rater has difficulty determining if an outcome deserves a 1 or 3 on writing, a 2 could be assigned. Table 1 includes a summary of *IFSP Rating Scale* scoring guidelines.

In previous studies, scores for IFSPs’ “writing” were treated as one item. Here however, writing was scored across two parts of the IFSP, namely present level of development and outcomes; thus, the instrument was comprised of 13 items in total. Because each item on the *IFSP Rating Scale* was applied multiple times on each IFSP (e.g., a writing score is generated for each outcome), the final item score for the IFSP was recorded as a mean, most often including a decimal. Prior to conducting the Rasch analysis, the data were recoded to whole numbers based upon the original anchoring of the scale, where 1, 3, and 5 served as the core. 1 and 5 were left as is, because these ratings were clearly defined as least and most. The middle scale score of 3 was assigned to any value falling in the

Download English Version:

<https://daneshyari.com/en/article/372609>

Download Persian Version:

<https://daneshyari.com/article/372609>

[Daneshyari.com](https://daneshyari.com)