# Using Mixed Methods to Establish the Social Validity of a Self-Report Assessment: An Illustration Using the Child Occupational Self-Assessment (COSA)

Jessica M. Kramer

*Journal of Mixed Methods Research* 2011 5: 52
DOI: 10.1177/1558689810386376

Journal of Mixed Methods Research
5(1) 52–76

DOI: 10.1177/1558689810386376
http://jmmr.sagepub.com

# Using Mixed Methods to Establish the Social Validity of a Self-Report Assessment: An Illustration Using the Child Occupational Self-Assessment (COSA)

**Jessica M. Kramer**[1]

## Abstract

Social validity is the extent to which procedures, goals, and outcomes are acceptable and important to end users and is not commonly addressed during assessment development. This article pilots a process of triangulating multiple methods to evaluate the social validity of a self-report assessment. These procedures were piloted using the Child Occupational Self-Assessment (COSA), a self-report of everyday activities for children with disabilities. This study collected COSA responses from 502 children with disabilities and conducted observations and interviews with 5 additional children with cerebral palsy. Data were aligned with social validity benchmark statements and integrated using data displays. Social validity was evaluated by examining congruencies and discrepancies in the results. Implications and the legitimacy of design decisions are discussed.

## Keywords

children with disabilities, self-report assessment, social validity

Assessment developers addressing validity are ultimately concerned with ensuring that inferences resulting from any assessment are valid, fair, and unbiased (Messick, 1995, 1998). For example, Messick (1995) urged assessment developers to consider the social consequences of test interpretation and to ensure that "adverse social consequences should not be attributable to any source of test invalidity" (p. 748). This concern represents the increasing influence of postpositivist epistemologies on the assessment development process. However, literature on assessment validity is typically limited to quantitative reports of construct validity, content validity, and concurrent validity, and does not address postpositivist conceptualizations of inference such as credibility, authenticity, or transferability (Onwuegbuzie & Johnson, 2006; Patton, 2002). New paradigms such as mixed methods provide an opportunity to incorporate these postpositivist conceptualizations of assessment validity into the assessment development process.

[1]Boston University, Sargent College of Health and Rehabilitation Sciences, Boston, MA, USA

**Corresponding Author:**
Jessica M. Kramer, Boston University, 635 Commonwealth Avenue (SAR 512), Boston, MA 02215, USA
Email: kramerj@bu.edu

The integration of multiple methods may have the capacity to provide new information and understandings of validity beyond those provided when individually examining the results stemming from either qualitative or quantitative methods. The purpose of this article is to respond to existing discussions in the assessment literature (Dellinger & Leech, 2007; Messick 1995; Onwuegbuzie & Johnson, 2006) by examining if and how multiple methods can be used to examine contemporary understandings of assessment validity. This article pilots a process of triangulating multiple data collection methods to evaluate an aspect of validity not typically evaluated in the process of assessment development: social validity. These procedures will be illustrated using data collected to develop the Child Occupational Self-Assessment (COSA), a self-report of everyday activities used in rehabilitation with children with disabilities.

The purpose of self-report assessments in rehabilitation is twofold: (a) to identify client needs and concerns in order to collaboratively determine the direction and content of intervention (American Occupational Therapy Association, 2002) and (b) to measure client-reported change after intervention to document effectiveness. Ensuring the validity of the interpretations made from self-reports to fulfill these two purposes requires researchers to answer different questions. The latter purpose requires researchers to ask questions regarding the stability of self-report assessments over time, and the sensitivity of a self-report assessment to capture change on a construct of interest. These types of questions are typically answered using psychometric studies. The former purpose requires research to investigate questions regarding the extent to which clients feel a self-report assessment is a vehicle by which they can identify and communicate concerns they wish to address during intervention. One way to approach these types of questions and evaluate if self-report responses can be interpreted in a credible manner is to evaluate social validity.

Social validity is the extent to which the procedures, goals, and outcomes of research are considered important and acceptable to the target population (Fawcett, 1991; Foster & Mash, 1999; Wolf, 1978). Researchers have traditionally applied the concept of social validity to studies investigating behavioral interventions by evaluating the importance and acceptability of the procedures, goals, and outcomes associated with therapeutic curriculum, techniques, and technologies (Foster & Mash, 1999). However, the growing use of self-report measures in rehabilitation (Hemmingsson, Kottorp, & Bernspång, 2004; King et al., 2004; Sturgess, Rodger, & Ozanne, 2002) and their dual purpose to measure outcomes and plan intervention points to the importance of establishing social validity during the assessment development process. That is, to ensure the credibility of intervention plans developed using self-reports, clients should find the procedures, goals, and outcomes of those self-report assessments important and acceptable.

The first step in applying the concept of social validity to assessment development is to consider how the facets of procedures, goals, and outcomes manifest in the self-report process. Self-report procedures encompass the processes by which individuals interact with questions and formulate responses. These responses may be open ended or provided by a rating scale. The amount of time this process takes and any emotional consequences of engaging in the process also are relevant when evaluating the social validity of procedures. These self-report procedures should be acceptable to the respondents and not considered upsetting, demeaning, or too lengthy. Second, the goal of self-reports is to identify respondents' needs and concerns in order to inform intervention planning. Therefore, self-reports should have the capacity to identify issues that are important and relevant to the target population of respondents. Third, the outcome of a self-report may be a score or other type of measure or index derived from responses to items. When social validity is applied to behavioral interventions, this third facet of outcomes typically refers to the extent to which the effects of the intervention are socially acceptable to end users. However, the underlying question driving the examination of this facet of social validity is, "Are consumers satisfied with the results?" (Wolf, 1978, p. 207). For a self-report, the outcome is an assessment of the "amount" of the underlying construct reported by respondents. This

assessment then informs intervention decisions. To ensure social validity, respondents should feel that the meaning others make through the interpretation of their scores corresponds with respondents' understanding of the underlying construct and the corresponding self-evaluation. This final facet of social validity most closely resembles well-accepted notions of construct validity. In fact, one should expect social validity to inform overall evaluations of construct validity. Messick (1995) points out that "the construct validity of score meaning is the integrating force that unifies validity issues into one unitary concept" (p. 748). Therefore, establishing the social validity of assessment outcomes, in addition to procedures and goals, can be conceptualized as one approach to better understand the meaning made from self-report assessments and, ultimately, to establish construct validity (Dellinger & Leech, 2007; Messick, 1998).

Examining a complex notion such as social validity may be particularly suited to a mixed methods approach. Mixed method designs use qualitative and quantitative research methods in combination to capitalize on each paradigm's epistemological stance, or different ways of knowing (Johnson & Onwuegbuzie, 2004). The use of complementary methods, each with alternate strengths, limitations, and ways of knowing, can lead to a more comprehensive understanding of the matter under investigation. Assessment developers recognize that using mixed methods can enhance the meaningfulness and usefulness of measurement instruments. Typically, researchers first employ qualitative research to better understand the construct of interest (see, e.g., Young, Rice, Dixon-Woods, Colver, & Parkinson, 2007). The qualitative findings then inform the development of assessment items. However, the validation phase may contain little or no qualitative research. With a few exceptions (Hitchcock et al., 2005; Lee & Greene, 2007; Myford & Mislevy, 1995), qualitative and quantitative data are rarely integrated to inform the assessment validation process. Similarly, social validity is typically established using multiple, but nonintegrated methods; normative comparisons are typically made using quantitative methods, and subjective evaluations are typically obtained by examining end users' perspectives through qualitative or quantitative reporting methods (see Foster & Mash, 1999, for examples). This article makes a unique contribution to the literature by piloting a process to integrate multiple forms of data, including measurement model data, open-ended responses, behavioral observations, and interview data, in a systematic manner to evaluate social validity. The purpose of this article is to pilot a process for triangulating and integrating multiple methods in order to evaluate the social validity of a self-report assessment and to determine the extent to which unique information is gained through this integration of multiple methods. The steps of this process are demonstrated in this article. The ramifications of the various design decisions made during this pilot process are reviewed to identify how multiple methods can be used to make legitimate decisions about the social validity of self-report assessments. The following methodological questions were encountered during this process and will be explored and discussed in this article:

1. What sampling technique is most credible when evaluating the social validity of self-report assessments?
2. How can researchers determine what evidence to use when evaluating the social validity of self-report assessments?
3. What techniques can integrate quantitative Rasch measurement data with qualitative data?

## Study Background

The process for integrating results was piloted using quantitative and qualitative data collected by this author to develop a self-report for children with disabilities: the COSA.

## The Child Occupational Self-Assessment

COSA (Keller, Kafkes, Basu, Federico, & Kielhofner, 2005) is designed to capture the perceptions of children with disabilities and their competence for performing everyday activities and the importance attached to those activities. Everyday activities are defined as frequently performed, goal-directed actions that children execute to take care of themselves, interact with others, learn, and play (American Occupational Therapy Association, 2002; World Health Organization, 2001). The COSA items range from basic everyday activities to more advanced activities.

The COSA includes 25 items that are rated using two 4-point scales: competence ("I have a big problem doing this," "I have a little problem doing this," "I do this ok," and "I am really good at doing this"), and importance ("Not really important to me," "Important to me," "Really important to me," and "Most important of all to me"). The COSA also includes a series of open-ended questions that invite children to share additional information. The open-ended questions include the following: "What are two other things you are really good at that we didn't talk about today," "What are two other things that you have a big problem with that we didn't talk about today," and "Is there anything else that is important to you that we didn't get to talk about—Would you like to tell me?" The rating scales and open-ended responses are used to identify client concerns that should be addressed in rehabilitation or areas of interest that can engage the child during intervention. The COSA does not generate a normative- or criterion- referenced score of competence or importance; rather, professionals administering the COSA are directed to identify gaps between the scale ratings. That is, those items for which the child reports low competence but high importance are considered a potential problem area that should be addressed during intervention.

The COSA was designed to be easy to use by children with a range of abilities. The scales are accompanied by visual symbols (faces and stars) to help signify differences between rating categories. Modifications can be made during administration to ensure that children can access and understand the assessment.[1] In addition to making modifications, professionals can administer the COSA in one of three ways: (a) an extended paper-and-pencil form, (b) a card sort, and (c) a summary matrix that does not include the symbols used in other forms.

## Overall Study Design

This study used a concurrent triangulation design, which enabled the researcher to corroborate findings across multiple methods to conduct a comprehensive evaluation of overall assessment validity (Creswell, Plano Clark, Gutmann, & Hanson, 2003). Each strand of the design addressed a method-specific research purpose. The researcher used quantitative methods to examine the psychometric properties of the COSA rating scales, such as rating scale function and item coherence. The researcher used qualitative methods to understand how children with disabilities constructed their sense of competence for everyday activities in the context of everyday life. A secondary aim of both strands was to gather some pilot information regarding the usefulness and acceptability of the COSA. The author sought to integrate and analyze the available qualitative and quantitative data in new ways to provide additional insights regarding the social validity of the COSA. Quantitative and qualitative data collection and analysis were concurrent; within each method, recruitment, procedure, and analyses were conducted in accordance with the methodological paradigm (Morse, 1998). Detailed descriptions of the methods used and the results stemming from each strand of this design have been reported elsewhere. However, this section will provide a brief summary of the qualitative and quantitative methods to better understand the nature of the data used in the mixed methods integration and analytical process.

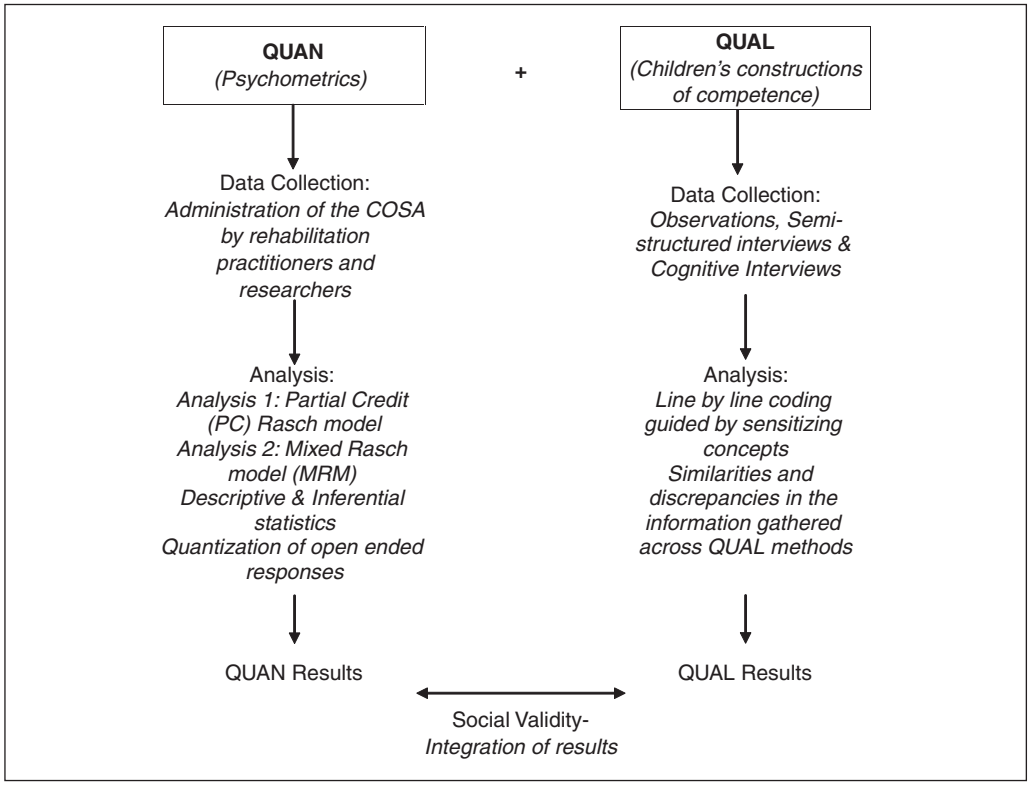


**Figure 1.** Concurrent triangulation study design detailing study phases

## *Quantitative Phase Procedures and Methods*

The researcher invited occupational and physical therapists to contribute COSA responses to a central database. Convenience sampling was used, and COSA responses from children meeting the following criteria were entered into the central database: children aged 6 to 17 years, children who completed the COSA in their primary language, and children with a diagnosed disability or receiving occupational therapy. This represents the targeted respondent group of the COSA. The database included COSAs gathered from the United States, the United Kingdom, Switzerland, Germany, and Iceland, and the assessments were administered in English, Italian, Icelandic, German, and British sign language. The method of administration and use of modifications were determined by each contributor. The final database included COSA responses from 502 children with disabilities (Table 1). This sample size was adequate for the quantitative analytical techniques applied.

Contributing therapists also reported information about the time needed to complete the COSA, modifications made during administration, and if results informed intervention planning. These questions were collected during the quantitative strand to provide information about the use of the COSA. These data were therapist reported. Although the literature includes instances in which social validity is evaluated by individuals other than the intended client group (Foster & Mash, 1999), it is difficult to draw credible conclusions regarding social validity from this information alone. Therapists' responses were analyzed using descriptive statistics.

To analyze the psychometric properties of the COSA, two quantitative analyses were conducted using children's responses, both based on the Rasch model (Wright & Stone, 1979). The

**Table 1.** Summary Demographics From QUANT Phase (*N* = 502)

| Demographics | *n* | Percentage |
|---|---|---|
| Gender | | |
| Male | 360 | 71.7 |
| Female | 142 | 28.3 |
| Ethnicity | | |
| Caucasian | 410 | 81.7 |
| African American/African | 35 | 7.0 |
| Hispanic | 22 | 4.4 |
| Other | 34 | 6.8 |
| Missing information | 1 | 0.2 |
| Major condition/diagnostic category | | |
| Developmental delay (e.g., autism) | 323 | 64.3 |
| Neurological (e.g., cerebral palsy) | 88 | 17.5 |
| Mental health | 46 | 9.2 |
| Other (e.g., chronic health condition) | 44 | 8.8 |
| Missing information | 1 | 0.2 |
| Practice setting | | |
| School (inclusive) setting | 233 | 46.4 |
| School (institutional/center) setting | 144 | 28.7 |
| Other setting (inpatient hospital, rehabilitation center) | 125 | 24.9 |
| Country | | |
| United States of America | 253 | 50.4 |
| United Kingdom | 187 | 37.3 |
| Other | 62 | 12.4 |
| Age | Years, Months (*n* = 497) | |
| Range | 6 years to 17 years 10 months | |
| Mean (*SD*) | 11 years 11.73 months (2 years 10.39 months) | |

Note: QUANT = quantitative.

Rasch model uses logarithms to estimate the location of individual test items, individual respondents, and rating scale categories on a common ruler representing more to less of the underlying construct. The depiction of items along this construct is referred to as an item hierarchy. The Rasch model also provides statistics, referred to as fit statistics, which indicate the extent to which each item belongs to the underlying continuum and each respondent could be scored in a reliable manner. When an item has acceptable fit, it suggests that respondents interpreted the meaning of the item in a similar and consistent (reliable) manner (Smith, Conrad, Chang, & Piazza, 2002). When a child has acceptable fit, his or her relative interpretation of the difficulty or importance of each item is consistent with the item hierarchy (Smith et al., 2002).

The first analysis applied the Rasch partial credit (PC) model (Masters, 1982); one analysis was conducted for each COSA scale. The second analysis applied the mixed Rasch model (MRM; Rost, 1990), which combines the Rasch model with latent class theory and identifies different groups of respondents, or classes for whom the Rasch model holds (i.e., item- and person-fit statistics are optimized within each class). In the MRM, separate groups of respondents may be identified by determining the personal and contextual variables that vary with class membership. As described later, the researcher used item hierarchies and fit indices generated during these analyses to evaluate social validity.

**Table 2.** Summary Demographics From QUAL Phase

| Name | Gender | Age | Disability Condition[a] | Adaptive Equipment | Ethnicity/Race |
|---|---|---|---|---|---|
| Patrick | Male | 7 years 11 months | Right hemiplegic cerebral palsy, Asperger's syndrome | Right leg brace | Multiracial |
| Lizzy | Female | 8 years 10 months | Right hemiplegic cerebral palsy | Right leg brace | Caucasian |
| Maria | Female | 10 years 4 months | Athetoid cerebral palsy | Bilateral leg braces, walker, wheelchair, augmentative communication device | Hispanic/Latina |
| Maki | Male | 11 years 4 months | Right-sided hemiplegia, secondary to in utero stroke | Right leg brace | Japanese American |
| Lenny[b] | Female | 17 years 1 month | Cerebral palsy | Wheelchair, augmentative communication device | "Don't know" |

Note: QUAL = quantitative.
a. Reported by parents
b. Reported by study participant.

The children's responses to the open-ended questions at the end of the COSA also provided data that could be used to evaluate the social validity. The author and a graduate research assistant used a consensus approach to code children's open-ended responses into thematic categories of activity areas such as organized sports and academics; the numbers of responses in each category were counted to generate frequencies for each activity area (Patton, 2002). These categories and frequency counts were then available for use in the process of evaluating social validity.

## *Qualitative Phase Procedures and Methods*

To better understand how children with disabilities constructed competence and to understand how these constructions may or may not be aligned with the COSA's construct of self-reported competence, the researcher initially sought to conduct a qualitative study with children aged 6 to 17 years having a range of disabilities (Patton, 2002). This was in keeping with the target population for the COSA. However, the first 3 children recruited to the study were children with cerebral palsy. The sampling strategy then shifted to gain an in-depth understanding of the competence of children with cerebral palsy (Patton, 2002). As data analysis was ongoing with data collection, recruitment continued until saturation was reached on themes related to the *primary* qualitative research question regarding children's constructions of competence for everyday activities (as reported elsewhere: Kramer & Hammel, IN PRESS). In total, 5 children participated in this study (Table 2).

Each child participated in a three-step procedure. The first step was a series of two participant observations at home and in the community. During the observation, the researcher and child took pictures of objects, other people, or the child engaged in different activities (Fasoli, 2003). The second step was a semistructured interview. Children viewed the photographs and elaborated on thoughts or feelings about the objects, tasks, or persons in the photographs with the goal of eliciting information about competence in the child's own words (Taylor & Bogdan, 1998). Children also sorted photographs into categories to indicate competence and importance of the activity using the rating categories and symbols from the COSA. These two steps were used to

elicit data that could answer the primary research question regarding children's construction of competence.

The third step was completing the COSA using cognitive interviewing techniques (Rebok et al., 2001). Cognitive interviews ask questions that reveal the processes that children use when responding to an assessment such as "what were you thinking about when you picked that response?" During this interview, children were also asked to reflect on what they liked and did not like about the process of completing the COSA. This third step was undertaken with the explicit purpose to gather information about the acceptability of the COSA. However, since this sample was limited to only five individuals with cerebral palsy, only limited conclusions could be drawn from the five cognitive interviews regarding social validity. However, integrating these data with the available quantitative data could potentially expand the conclusions that could be drawn regarding social validity.

For all three steps, the researcher completed detailed field notes after each observation. All interviews were video recorded and transcribed, and picture rankings and COSA responses were recorded on data collection forms. Summary photo albums that included initial interpretations of observation and interview data were generated to member check data; however, only one parent responded to the researcher after receiving the photo album to indicate her and her child's approval of the material.

To analyze the qualitative data in a way that could inform an evaluation of social validity, the researcher used the sensitizing concepts of "the importance of the COSA" and the "acceptability of the COSA" to guide line by line coding of each cognitive interview transcript (Patton, 2002). These sensitizing concepts helped the researcher identify data relevant to social validity. Codes were also created to describe children's responses to and opinions of the COSA. As each new transcript was coded, codes were compared with codes generated from the analysis of other transcripts (Miles & Huberman, 1994). Codes were collapsed and expanded to generate a final list of children's opinions regarding the procedures, goals, and outcomes of the COSA (Patton, 2002). The researcher then looked for ways in which data collected in the other steps of the qualitative study could be analyzed to provide information about the acceptability and importance of the procedures, goals, and outcomes of the COSA. To do this, data from each child were reviewed to identify similarities or discrepancies in the information gathered during observations, reported in the open-ended interviews, and during the COSA cognitive interviews. The researcher tracked examples of these situations for each child using a table, consisting of the columns *similarities* and *discrepancies*, and the examples were available for use in the process of evaluating social validity.

## Evaluation of Social Validity: Analytical Procedures

Several scholars have cited the need to make explicit the reasoning behind method integration (Greene, 2008; McConney, Rudd, & Ayres, 2002). Greene (2008) asked, "Around what does the mixing happen?" (p. 17). In this pilot, the three facets of social validity drove the mixing and integration of quantitative and qualitative results. First, the researcher created a benchmark statement that would represent acceptable social validity for the COSA. These benchmark statements were informed by two sources: the literature describing the evaluation of social validity of intervention research (Fawcett, 1991; Foster & Mash, 1999; Wolf, 1978), and the literature on children's advocacy and involvement in intervention planning (Betz, Redcay, & Tan, 2003; Cavet & Sloper, 2004; Gan, Campbell, Snider, Cohen, & Hubbard, 2008; Turner, 2003; Valnes, Hare, Block, Branch, & Broshears, 2003). Second, the researcher identified the available quantitative (QUANT) and qualitative (QUAL) evidence that could provide information relevant to that benchmark. The relevant results could then be integrated to evaluate each facet of social validity.

**Table 3.** Facets of Social Validity and Evidence Used to Evaluate Social Validity Benchmarks

| Facet of Social Validity | Benchmark | Evidence Used to Evaluate Benchmark |
|---|---|---|
| Procedures | The process of completing the COSA is enjoyable and not too difficult for children. | • QUAL: Children's verbal and behavioral responses while completing the COSA.<br>• QUAL: Children's reactions to the structure and presentation of the COSA.<br>• QUANT: Time needed to administer the COSA. |
| Goals | Children can use the COSA to identify important and relevant interests, needs, or concerns. | • QUAL: Activities identified by children in observation can be matched to COSA items, and conversely, children relate COSA items to everyday life experiences.<br>• QUANT: Therapists' reports regarding usefulness of COSA to identify goals for intervention.<br>• QUAL: Children's desire to use the COSA with their therapist to identify rehabilitation goals.<br>• QUAL + QUANT: Congruency between children's responses to the open-ended questions in the QUANT study and activities shared by children during the QUAL study. |
| Outcomes | Interpretations made from children's responses correspond with their understanding of the underlying construct and the corresponding self evaluation. | • QUAL: Rating scale use<br>  • COSA item rankings match children's reports during QUAL observation.<br>  • Children interpret the rating scale categories in the intended manner.<br>• QUANT: Rating scale use<br>  • All children use the 4-point rating scale in the intended manner.<br>• QUAL + QUANT: Consistent, intended interpretation of items by children across the methods.<br>• QUAL + QUANT: Congruency of the relative difficulty and importance of everyday activities within and across the methods. |

Note: QUAL = quantitative; QUANT = quantitative.

This approach offers an alternative to the typical procedures used to evaluate social validity. In most instances, social validity is assessed by asking research participants to provide a subjective evaluation of various aspects of the acceptability and importance of the intervention through the use of survey or open-ended questions about the procedures, goals, and outcomes (Fawcett, 1991; Foster & Mash, 1999). In this study, existing data were triangulated and compared against the benchmark statement, and social validity can be inferred based on the extent to which multiple forms of data coalesce in support of the benchmark statement. This process actively inserts postpositivist approaches to generating knowledge, notably the qualitative approach of making constant comparisons between multiple sources of data to identify commonalities as well as outliers (Strauss & Corbin, 1998), into the assessment development and social validation process. Table 3 depicts the final outcome of this process.

In some instances, qualitative and quantitative results each independently contributed to the evaluation of a facet of social validity. For example, the time required to administer the COSA (QUANT) and children's nonverbal behaviors (QUAL) were two results used to evaluate the social validity of procedures. In instances such as this, to evaluate social validity, the researcher

gave equal weight to each result, identified commonalities or discrepancies among those results, and determined the extent to which the results in combination endorsed the benchmark statement. This type of mixing is typical of triangulation designs, which mix at the level of inference (Greene, 2008).

In other instances, qualitative and quantitative results could be integrated and analyzed in new ways to evaluate social validity. In these situations, the researcher created several types of integrated data displays, as described below (Lee & Greene, 2007). This process enables results arising from the quantitative study to be mixed with results emerging from the qualitative study at the analytical level, opening up new avenues of mixed methods practice (Greene, 2008).

One data display evaluated the congruency of information shared across methods by children with disabilities (see Table 4). This data display required a "quantization" of open-ended responses reported on the COSA in the QUANT study, as previously described. The researcher compared the quantitized frequency counts from the content analysis of activity areas and representative examples of open-ended responses given by children in the QUANT study with quotes obtained during the QUAL study. Congruency between activities identified using the COSA in the QUANT study and in "real-life" contexts in the QUAL study provides evidence supporting the social validity benchmark for the goal of the COSA.

A second data display evaluated the relative competence and importance reported for everyday activities across methods (see Figures 2 and 3). This data display compared the item hierarchies from the QUANT analyses with data from the QUAL cognitive interview. For each child in the QUAL study, the items were ranked based on the response category selected for each item. Each child's item rankings were then systematically compared with each Rasch item hierarchy. Similarities and differences in the relative competence and importance reported for everyday activities across the methods provide a better understanding of the competence and importance children with disabilities report for everyday activities. The extent to which the relative difficulty and importance of everyday activities are congruent within and across the methods indicates if individual versus universal interpretations of COSA responses lead to a socially valid outcome.

A third data display used extreme case analysis to investigate the consistency of item interpretation across methods (Lee & Greene, 2007; see Table 5). Items with poor fit to the Rasch model in the QUANT study were one type of extreme case. Items for which children's descriptions in the QUAL cognitive interviews differed from the intended item meaning as presented in the COSA administration manual were a second type of extreme case. Items interpreted in an inconsistent manner (extreme cases) by children in both the QUAL and QUANT studies would be considered threats to social validity of the outcome of the COSA. This information also helps determine if the current approach to interpreting the COSA responses at an individual level leads to an endorsement of the outcomes benchmark.

## Results of Mixed Method Data Integration

This results section provides a summary of the pilot data integration and analysis process to evaluate social validity.

### *Social Validity: Procedures*

Children in the QUAL study demonstrated their interest and engagement in the COSA with the following behaviors: smiling when reading an item, thinking behaviors such as lifting their eyes upward, and asking questions about items. Children in the QUAL study reported a range of favorite things about the assessment, including the symbols, items, and open-ended questions.

All children reported that they felt comfortable with the symbols used on the COSA. Children gave the following advice about completing the COSA:

> Relax . . . it's not really a big competition or something, and it's not just for nothing. It's for something, but you can relax too. (Lizzy)
>
> Think about it [before picking the answers]. (Lenny)
>
> It's easy. (Maria)

The children in the QUAL study also occasionally demonstrated the following behaviors indicative of boredom or frustration: sighing, distraction (such as looking out the window or *spacing out*), and becoming impatient (responding with "I don't know" or quickly selecting responses to a series of items). Patrick and Lizzy expressed frustration with the length of the assessment, and Patrick said, "It's too hard." As the cognitive interview took approximately 60 minutes and was completed in one session, it is understandable that children would fatigue, especially younger children. In the QUANT study, the COSA took an average of 27 minutes to complete, and administration times ranged from 5 minutes to 2 hours 15 minutes. However, longer administration times occurred when the COSA was administered over several sessions. This modification strategy may minimize fatigue and frustration by reducing the amount of information presented at one time.

The manner in which children handled their own errors also revealed concerns with the administration directions. For example, after Lizzy accidentally circled a rating response on the wrong line, when it came time to rate the item on that line, she did not erase and re-rate her response. In another example, even though the directions specified that answers could be changed at any time, Maria asked the researcher for permission to change a response. These data suggest that the directions as currently written do not adequately communicate to children that it is ok to change responses or correct any "mistakes."

### *Social Validity: Goal*

In the QUAL study, children identified a range of activities that were important, that they were good at doing, or that they had a hard time doing. These activities were in the areas of self-care (i.e., tying shoes, showering, cleaning room), school (i.e., doing homework, following rules), social interactions (i.e., parents, siblings, friends), and leisure activities (i.e., playing with pets, videogames, art, physical activities). Items representative of each of these areas are included on the COSA.

Children in the QUAL study also related COSA items to everyday life experiences. For example, during his observation, Patrick mentioned that he used reading as a strategy to make his "engine just right." Later, in response to the COSA item "Calm myself down when I am upset," Patrick again shared that he read a book. When Lenny rated the COSA item "Think of ways to do things when I have a problem," she told a story about how she got around the second floor of her school after the elevator broke.

Professionals' reports from the QUANT study indicate that the COSA provides information relevant to intervention planning. The majority (80.2%) of responding therapists planned to use the information from the COSA to create intervention goals. Although these findings appear to endorse the usefulness of the COSA, the professionals' reports conflict with children's opinions of the COSA obtained during the QUAL study. Only Lenny, the oldest, indicated that she would consider using the COSA with a professional to set intervention goals. Two other children explicitly stated that they would not like to use the COSA to identify their goals for intervention because they did not want to spend more time attending therapy.

**Table 4.** Activity Areas Reported Across QUAL and QUANT Phases

| | QUANT Study: Open-Ended COSA Responses | | | | |
|---|---|---|---|---|---|
| | Frequency of Responses For | | | | |
| Activity Area | Things I am good at doing | Things I have a problem doing | Things that are important to me | Example of Specific Written Response | QUAL Study Example of Child Quote |
| Organized sports | 227 | 22 | 18 | Baseball<br>Horseback riding<br>Bowling | "Play [baseball] at the field . . . it's the wolverines. This was our last game and we won championship."—Patrick<br>"I've been on swim team since I was six, and I'm almost 9 now."—Lizzy |
| Games and playing | 45 | 2 | 0 | Playing with siblings/ friends<br>Dress up<br>Chess | "Sometimes I seem to be better and sometimes I just get to tap [the ping pong ball] once, twice, three times and then [makes falling motion] . . ."—Maki<br>"[The ball pit] is fun—I can swim, jump off, do a dive."—Patrick. |
| Computer and video games | 84 | 4 | 11 | Nintendo | "I have more time [to play] my gameboy . . . I've been doing it one year. I wanted to get it."—Maria<br>"[This game character] is so good."—Patrick |
| Cognition | 6 | 36 | 0 | Having ideas<br>Concentrating<br>Remembering | "I go off the world . . . [when I am] bored."—Lenny |
| Home tasks and chores | 54 | 28 | 28 | Cooking<br>Taking care of pets<br>Babysit | "In the summer, I often water [the garden]."—Maki<br>"Put things in order . . . all the books in one place."—Maria |
| Self-care | 16 | 15 | 2 | Brushing teeth | "Sooner or later I'll have to do it, 'cause not all shoes are Velcro . . . and then when we're in a rush [my mom] just ties [my shoes]."—Maki<br>Maria reported a hard time buttoning her pants. |
| Gross motor and mobility | 29 | 42 | 0 | Walking<br>Jump rope | "If I hold the camera, I might fall. I might [get] cut a little."—Patrick<br>"I was in first grade, and in gym we did a unit about jumping rope, so I got better then."—Lizzy |
| Arts and crafts | 71 | 11 | 5 | Models<br>Drawing<br>Taking pictures | "The feel of [clay] . . . it feels good in my hands . . . I had art classes."—Lenny<br>"I was surprised I'm that good [at painting] . . . you need a lot of wrist to get the tip of the leaf."—Maki |

*(continued)*

**Table 4.** (continued)

| Activity Area | QUANT Study: Open-Ended COSA Responses | | | | QUAL Study |
|---|---|---|---|---|---|
| | Frequency of Responses For | | | Example of Specific Written Response | Example of Child Quote |
| | Things I am good at doing | Things I have a problem doing | Things that are important to me | | |
| Performance arts | 42 | 11 | 9 | Playing an instrument<br>Dancing | "Singing [in my heart]."—Lenny |
| School and academics | 99 | 152 | 34 | History<br>Writing | Maria reported winning first place in her class science fair.<br>"Writing papers [and communicating ideas to my aid] . . . is harder."—Lenny<br>"Well, math is hard for me, and sometimes I'm just stuck on this one problem, I can't figure out any ways to do it, my teacher can't help me 'cause it's a test."—Lizzy |
| Specific skills, talents, and interests | 35 | 0 | 20 | "Copying accents"<br>"Dealing cards"<br>Motorbikes | "Animals . . . there are so many different kinds that have different shapes and make different sounds."—Patrick<br>"Church is important because "I love our God."—Lenny |
| Social/interpersonal | 55 | 70 | 32 | Being/making friends<br>Helping others<br>Not taking things personally | "My friends at school are great . . . we go out on weekends."—Lenny.<br>"Sometimes when my friends really wanna do things I just give in—I just say 'let's go do what you want to do—I don't really care.'"—Lizzy |
| Family | 10 | 32 | 48 | Baby brother<br>Mom and dad | "I am a good sister because I be nice but sometimes I get mad . . . [when my brother] eats all my chips!"—Maria<br>"One night, me and my brother and my two sisters and my dad were on the trampoline all together and we were just jumping and we were throwing the balls out . . . it was really fun."—Lizzy |
| Managing behaviors | 0 | 52 | 14 | Getting upset<br>Not swearing | "I was first scared, but when I did it I was normal. Just right engine."—Patrick<br>"My teacher in 2nd grade has this green light, a yellow light, and a red light . . . if you do something bad you get a warning and . . . you move to green. . . . I've never even gotten to green, once!"—Lizzy |

Note: QUAL = quantitative; QUANT = quantitative.

| Lenny's COSA Ratings* | | MRM Analysis: Importance Class 1 Item Hierarchy | |
|---|---|---|---|
| Rating Category | COSA Item | Item | Item Estimate (Error)** |
| Important to me | • Keep my body clean<br>• Buy something myself<br>• Get enough sleep<br>• Ask my teacher questions<br>• Make others understand my ideas<br>• Calm myself down when I get upset<br>• Make my body do what I want it to do<br>• Use my hands to work with things<br>• Finish what I am doing without getting tired too soon | Do things with my classmates<br>Buy something myself<br>Get my chores done<br>Ask my teacher questions<br>Make others understand my ideas<br>Keep working on something even when it gets hard<br>Think of ways to do things when I have a problem<br>Keep my mind on what I am doing<br>Choose things that I want to do | 0.51 (.12)<br>0.47 (.11)<br>0.41 (.11)<br>0.26 (.11)<br>0.25 (.11)<br>0.21 (.11)<br>0.14 (.11)<br>0.12 (.11)<br>0.10 (.11) |
| Very important to me | • Keep working on something even when it gets hard<br>• Finish my work in class on time<br>• Follow classroom rules<br>• Do things with my classmates<br>• Do things with my friends<br>• Do things with my family<br>• Keep my mind on what I am doing<br>• Choose things that I want to do<br>• Have enough time to do things I like<br>• Get my chores done<br>• Eat my meals without any help | Calm myself down when I am upset<br>Finish what I am doing without getting tired too soon<br>Have enough time to do things I like<br>Finish my work in class on time<br>Get my homework done<br>Follow classroom rules<br>Get around from one place to another<br>Do things with my friends<br>Take care of my things<br>Use my hands to work with things<br>Eat my meals without any help | 0.06 (.11)<br>0.05 (.11)<br>0.04 (.11)<br>0.02 (.11)<br>-0.01 (.11)<br>-0.05 (.11)<br>-0.08 (.11)<br>-0.09 (.11)<br>-0.17 (.11)<br>-0.21 (.11)<br>-0.23 (.11) |
| Most important of all to me | • Take care of my things<br>• Get around from one place to another<br>• Get my homework done<br>• Think of ways to do things when I have a problem<br>• Dress myself | Make my body do what I want it to do<br>Get enough sleep<br>Dress myself<br>Keep my body clean<br>Do things with my family | -0.27 (.11)<br>-0.28 (.11)<br>-0.31 (.11)<br>-0.39 (.11)<br>-0.56 (.11) |

**Figure 2.** Lenny's COSA importance ratings compared with importance item hierarchy from Analysis 2: MRM, Importance Class 1.

Note: COSA = Child Occupational Self-Assessment; MRM =mixed Rasch model.

*Lenny did not use the lowest values rating category "Not really important to me." **Given in logits, an interval-level estimate. Higher estimates indicate less importance, and lower estimates indicate more importance.

Children's open-ended responses in the QUANT study represented a variety of activity areas (Table 4). Children in the QUAL study shared similar interests and concerns. Organized sports and school/academics were activity areas most frequently brought up by children in the QUANT study. All activity areas were mentioned across methods.

| Maki's COSA Ratings* | | PC Analysis: Competence Item Hierarchy | |
|---|---|---|---|
| Rating Category | COSA Item | Item | Item Estimate (Error)** |
| I have a little problem doing this | • Buy something myself<br>• Choose things that I want to do<br>• Calm myself down when I am upset | Keep my mind on what I am doing | 0.70 (.05) |
| | | Get enough sleep | 0.60 (.05) |
| I do this ok | • Get around from one place to another<br>• Do things with my classmates<br>• Finish my work in class on time<br>• Make others understand my ideas<br>• Think of ways to do things when I have a problem<br>• Keep working on something even when it gets hard<br>• Make my body do what I want it to do<br>• Use my hands to work with things<br>• Finish what I am doing without getting tired too soon | Calm myself down when I am upset | 0.48 (.05) |
| | | Think of ways to do things when I have a problem | 0.47 (.05) |
| | | Get my homework done | 0.43 (.05) |
| | | Keep working on something even when it gets hard | 0.35 (.05) |
| | | Get my chores done | 0.35 (.05) |
| | | Make others understand my ideas | 0.28 (.05) |
| | | Finish what I am doing without getting tired too soon | 0.22 (.05) |
| | | Finish my work in class on time | 0.21 (.05) |
| | | Follow classroom rules | 0.16 (.05) |
| | | Buy something myself | 0.15 (.05) |
| | | Make my body do what I want it to do | 0.14 (.05) |
| | | Ask my teacher questions when I need to | 0.04 (.06) |
| I am really good at doing this | • Ask my teacher questions when I need to<br>• Get my homework done<br>• Follow classroom rules<br>• Do things with my friends<br>• Do things with my family<br>• Keep my mind on what I am doing<br>• Have enough time to do things I like<br>• Take care of my things<br>• Eat my meals without any help<br>• Get enough sleep<br>• Get my chores done<br>• Dress myself<br>• Keep my body clean | Take care of my things | -0.09 (.06) |
| | | Do things with my classmates | -0.23 (.06) |
| | | Have enough time to do things I like | -0.26 (.05) |
| | | Choose things that I want to do | -0.34 (.06) |
| | | Use my hands to work with things | -0.35 (.06) |
| | | Keep my body clean | -0.43 (.06) |
| | | Get around from one place to another | -0.49 (.06) |
| | | Eat my meals without any help | -0.50 (.07) |
| | | Do thing with my family | -0.52 (.06) |
| | | Dress myself | -0.63 (.06) |
| | | Do things with my friends | -0.71 (.06) |

**Figure 3.** Maki's COSA competence ratings compared with competence item hierarchy from Analysis 1: PC.

Note: COSA = Child Occupational Self-Assessment; PC =partial credit.

*Maki did not use the lowest competence rating category "I have a big problem doing this." **Given in logits, an interval-level estimate. Higher estimates indicate less competence, or more difficulty. Lower estimates indicate higher competence.

## Social Validity: Outcome

The congruency between information generated in QUAL observations and COSA item rankings varied. Sometimes children selected a COSA response that matched the information shared during the observation. For example, during her observation, Maria shared that she had a hard time buttoning her pants. Later, when completing the COSA, she rated that she had a little problem with the item "dress myself." However, sometimes children's COSA reports contradicted information gathered in previous stages of the QUAL study. During the photo ranking, Patrick indicated that he had a "big problem" putting on his vest, and during observations he reported that it was hard to put on his shoes. However, when completing the COSA, he reported that he was "really good" at getting dressed. The congruency between children's reports using the COSA and through interactions embedded in real-life contexts did not appear to be dependent on the child's age or disability severity, as all children in the QUAL study had COSA responses that both contradicted and were aligned with the information shared during observations.

Children in the QUAL study appeared to use the symbols accompanying the rating scales to discriminate between the different rating categories as Lizzy explained, "The one smiley face shows you do it ok, you're not super good at it, you don't have a problem doing it." However, in some instances, children's interpretations of the rating scales did not match the intended meaning. For example, Patrick associated how much he liked something with the importance categories, although enjoyment is not a condition for importance.

The QUANT MRM analysis of the rating scale determined that approximately 50% of children used the COSA rating scales in the intended manner, as a 4-point continuum of less to more competence and importance (Kramer, Smith, & Kielhofner, 2009).[2] The remainder responded in an unintended manner as their interpretation of both rating scales resembled a 2-point scale;[3] that is, they were unable to make meaningful distinctions between the lower three rating categories. These children were more likely to be younger and have intellectual disabilities. This suggests that some children bifurcate the response categories to report whether they have more or less competence or importance for any particular activity.

Children's interpretation of the COSA items appeared relatively consistent. In the QUANT study, only two items were identified as extreme cases using the poor fit criteria. However, additional items were identified as extreme cases in the QUAL study, as indicated in Table 5. Only the item "Make my body do what I want it to do" emerged as an extreme case in both the QUANT and QUAL studies.

The relative competence and importance reported for everyday activities varied within methods: within the QUANT MRM analysis, the two classes of respondents produced different item hierarchies; in the QUAL study no two children had the same item rankings. Relative competence and importance for everyday activities also varied across methods; that is, no Rasch item hierarchy was an exact match to the rankings obtained during the QUAL study. An example is the data display in Figure 2, which compares Lenny's importance ratings and an item hierarchy derived from QUANT MRM analysis, Importance scale, Class 1; this item hierarchy was determined to be most congruent with Lenny's responses. Some items shared similar rankings; for example, both Lenny and Importance Class 1 reported relatively less importance for the item "Make others understand my ideas" and both indicated that "Dress myself" was one of the most important activities. However, Lenny differed from Importance Class 1 in that "Think of ways to do things when I have a problem" and "Get my homework done" were two of the most important activities, unlike Importance Class 1.

Other instances of congruencies and discrepancies between the relative importance reported for items across methods include the following:

- Four out of the five QUAL children reported that thinking of ways to do things when they had a problem was "most important" or "very important" to them, and relatively high importance was reported by children in the QUANT PC analysis.
- Four out of five QUAL children indicated that it was "most important" to get their homework done, whereas children in the QUANT study assigned relatively low importance to this item as given by the PC analysis.
- Four out of five QUAL children used one of the two highest rating categories to indicate the importance of "Take care of my things." Similarly, the PC analysis indicates that children in the QUANT study assigned the most importance to this item.

As an example of relative difference in reported competence across methods, Figure 3 compares Maki's competence ratings with the competence item hierarchy from the QUANT PC analysis; this was determined to be most congruent with his responses. Like the children in the Partial Credit analysis, Maki reported that he did self-care items "really good." Both Maki and children in the QUANT sample reported lower competence for the item "Calm myself down when I am upset." However, children's reports of competence in the QUANT study were in direct converse to Maki's indication that he was really good at getting his chores done, following classroom rules, and getting his homework done.

Other instances of congruencies and discrepancies between the relative competence reported for items across methods are as follows:

- Four out of the five QUAL children reported that they were "really good" at following classroom rules, and in the QUANT PC analysis, children reported moderate competence for this item.
- Three out of five QUAL children reported that they were "really good" at keeping their mind on what they were doing, and two reported they did this "ok," whereas children in the QUANT study indicated this was the most difficult item as given by the PC analysis.
- Three out of five QUAL children reported that they were "ok" at using their hands, and one reported a "little problem," but children indicated relatively high competence for this item in the QUANT study.

These findings suggest that the relative competence and importance assigned to COSA items varies across individual children, as well as groups of children.

## Discussion

This article piloted a process whereby multiple data collection methods were used to evaluate social validity. Social validity is a complex construct that has been traditionally assessed by soliciting the perspectives of end users and other informants such as professionals using either quantitative or qualitative techniques (Foster & Mash, 1999). The process used in this article to integrate available qualitative and quantitative data provided a more credible evaluation of social validity than would be gained by examining social validity using one method only. Several examples in this article illustrate how the use of multiple methods provided a comprehensive evaluation of social validity. When evaluating the social validity of procedures, the data gathered in the QUAL phase revealed that the children were hesitant to change their responses, yet this information was not available in the QUANT data collected. As another example, the process of integrating QUANT Rasch item hierarchies and QUAL item rankings revealed the relative competence and importance for COSA items varied across individual children as well as groups of

**Table 5.** Extreme Case COSA Items and Comparisons Across QUAL and QUANT Phases

| | | | QUANT:COSA Item Fit | | |
|---|---|---|---|---|---|
| COSA Item | Extreme Case Inclusion | Definition in COSA Manual | Analysis and Scale | Fit Status[a] | QUAL: Children's Descriptions |
| Get my chores done | Poor fit, PC analysis | You are able to finish jobs asked of you without any help from others. Chores could involve jobs assigned in the home, classroom, or other setting as appropriate. | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Unacceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | "My books—put them on the shelf . . . knickknacks—like my monkey bank is a knickknack…[so I should] put it on the shelf."—Patrick<br>"Sometimes I don't do them even when I supposed to, 'cause I forget or I just really aren't in the mood for chores . . . even though I still have to do them!"—Lizzy |
| Get around from one place to another | Cognitive interview | You can move your entire body to get to where you need to go. (This item refers to mobility within the client's various environments rather than gross motor skills.) | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | "I can't walk from the place . . . it's too far."—Patrick<br>"Get to a friend's house by yourself . . . I don't usually do that unless I'm going to play with my friend across the street."—Lizzy<br>"My room to my sister's room . . . all the way down there."—Maria<br>"Wheelchair . . . Buildings . . . Getting in."—Lenny<br>"My mom always drives me to school and everywhere—she's pretty much always early. Instead of my dad . . . he's on time, or he's late."—Maki |
| Buy something myself | Cognitive interview | You are able choose an item to buy and know how much money to give to a cashier. If you had a dollar, you would know how to purchase a needed or desired item. The purchase could take place within school, the community, or any other setting. | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | ". . . I can like give the cashier the money and stuff, but sometimes I need my mom's help . . . my mom has to do the credit card and type in the numbers. I can't do that part. 'Cause I don't have a credit card."—Lizzy<br>"I could, but lots of the time, my mom does it instead of me."—Maki |

*(continued)*

**Table 5.** (continued)

| COSA Item | Extreme Case Inclusion | Definition in COSA Manual | QUANT:COSA Item Fit: Analysis and Scale | QUANT:COSA Item Fit: Fit Status[a] | QUAL: Children's Descriptions |
|---|---|---|---|---|---|
| Get enough sleep | Cognitive interview | You sleep enough so that you have the energy to do the things you need or want to do. | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | "I go to bed at like 8:30 . . . and then getting up at 7:15 in the morning . . . So I have enough time to get ready for school at 8:15 . . . 'Cause we're in school saying the pledge of allegiance at 8:15."—Lizzy<br>"I didn't sleep last night."—Maria |
| Make my body do what I want it to do | Poor fit, MRM analysis; cognitive interview | You can make your body move to play, work and do the things you want to do. (This item refers to gross motor skills.) | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Unacceptable<br>Acceptable<br>Acceptable<br>Acceptable | "I didn't really know what I was thinking."—Lizzy<br>". . . because my hand doesn't work as well."—Maki |
| Do things with my classmates | Cognitive interview | You are able to get along with the children in your class to work and share in school. | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | ". . . sometimes I get sidetracked and I start thinking about other things."—Maki<br>"Sometimes it's hard to work with a group . . . if you're with some of the not as well behaving kids . . . they don't pay attention—you need their help, but they don't pay attention."—Lizzy |
| Use my hands to work with things | Cognitive interview | You can make your hands and fingers move to do things with games, school supplies, or other objects. | PC: Competence<br>PC: Importance<br>MRM: Competence Class 1<br>MRM: Competence Class 2<br>MRM: Importance Class 1<br>MRM: Importance Class 2 | Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable<br>Acceptable | "I was thinking . . . this is kind of a trick question . . . I don't know if you mean, both my hands, or just one hand, or just the other hand . . ."—Lizzy<br>"Left hand . . . holding a camera."—Maria |

Note: QUAL = quantitative; QUANT = quantitative; PC = partial credit; MRM = mixed Rasch model.

a. PC analysis: Unacceptable fit to the model when MnSq ≥ 1.4 is associated with Zstd ³ 2.0 (Wright & Linacre, 1994). MRM analysis: Unacceptable fit to the model when Q-index > 0.3 and/or Zstd Q-index significant at $p < .05$ or $p > .95$ level (Wagner-Menghin, 2006).

children with similar functional limitations. The integration process was not able to reveal how or why this occurred because of the limited data available for analysis, but the integration process enabled the researcher to identify additional concerns that must be addressed in order to enhance the social validity of the COSA. Overall, these findings highlight the benefits of using and integrating multiple methods to evaluate social validity of self-report assessments.

However, several aspects of this study design require critical evaluation to determine the legitimacy of the procedures used to establish social validity of self-report assessments (Onwuegbuzie & Johnson, 2006). The following issues may be of particular interest to the mixed methods research community: the integration of samples, the selection of multiple methods, and the conversion of data for integration.

The primary legitimacy issue is in regard to sample integration. Mixed methods scholars have noted that concurrent mixed methods designs may use a variety of sampling techniques, and that the concurrent samples may be identical, parallel, nested, or multilevel (Onwuegbuzie & Collins, 2007; Teddlie & Yu, 2007). In this study, both the QUANT and QUAL study relied on convenience sampling methods. However, the broad inclusion criteria used in the QUANT study resulted in a highly heterogeneous sample of international children with disabilities. Conversely, the QUAL sample consisted only of children with cerebral palsy from one geographical location. The researcher attempted to create parallel samples (Onwuegbuzie & Collins, 2007) by using maximum variation sampling to obtain maximum heterogeneity in the QUAL sample; this attempt was not successful. Although different on disability composition and nationality, the two samples were similar in several other regards, including representation of a range of ethnic groups and age range. However, the samples were not drawn from, nor do they represent, the same population (Onwuegbuzie & Johnson, 2006; Teddlie & Yu, 2007). One could consider that children in the QUAL sample represent a theoretical subgroup of the children in the QUANT sample, which included children with cerebral palsy. Triangulating results from a subgroup to a population in which that subgroup belongs is similar to a nested sampling strategy (Onwuegbuzie & Collins, 2007); except in this study, the subgroup was not recruited directly from the larger population due to practical and logistical restrictions. The legitimacy of this sampling design hinges on the following questions: Can data from a theoretical subsample be integrated with data from a second, broader sample? In keeping with the pragmatic philosophy underlying mixed methods research, is it possible that a theoretical subgroup sampling strategy is appropriate for some research questions about social validity?

For this particular investigation of the social validity of the COSA, the answer may be no. If the research question was to evaluate the social validity of the COSA with all target respondents (children with a range of disabilities aged 6-17), then both QUAL and QUANT samples should represent that target population of respondents. However, if the research question was to identify if social validity of the COSA varied across potential target respondents, then the use of a theoretical subgroup sampling strategy may be appropriate.

Generating evidence for all validity, including social validity, is an ongoing process (Messick, 1995). Information generated through the systematic application of an assessment with multiple groups in multiple contexts builds evidence for social validity over time. This study used the COSA with two different samples, both of which consisted of individuals to whom the assessment is targeted. Therefore, even in the presence of this study's sampling limitations, some information regarding the social validity of the COSA can be gained. The similarities in the types of activities identified across QUANT and QUAL samples provides initial evidence that the COSA has the capacity to identify important and relevant interests, needs, and concerns. Findings also suggest that children's perceived competence for and the importance of everyday activities are highly individualized. Differences in item ordering within and across methods suggest that the relative competence and importance of items may vary according to each child's unique

circumstances and context. Findings stemming from this mixed methods evaluation point to important areas for future research. Replicating the qualitative data collection methods with a broader range of respondents, and integrating those results with the QUANT results would enable the researcher to determine if similar inferences regarding social validity of the COSA can be made with other disability groups.

Assessment developers seeking to examine the social validity of self-reports using mixed methods should consider several sampling guidelines: (a) The nature of the research question, (b) the group to which an assessment is targeted, and (c) the extent to which the assessment outcome is hypothesized to differ among subgroups of respondents. First, in keeping with the pragmatics underlying the mixed methods research paradigm (Johnson & Onwuegbuzie, 2004), sampling strategies should be guided by the research question. Second, regardless of the sampling strategy used, all samples should reflect the target population of respondents. Third, if the research question supports the use of a theoretical subgroup sampling strategy, the theoretical subgroup sampling strategy may only be appropriate if there is not an expected relationship between subgroup characteristics and the construct measured by the self-report, or if the relationship is unknown. To give an example from this study, it is possible that the data generated by the QUAL sample reflect the unique lived experience of children with cerebral palsy, and therefore may differ from the information gathered from the larger, heterogeneous group of children with a range of disabilities. This potential relationship calls into question the inferences made from this study's integration of data from samples drawn from different populations. Further research is needed to confirm this relationship, which could then lead to a more appropriate sampling strategy to evaluate the social validity of the COSA.

The second legitimacy issue is the benchmarking process used to select multiple forms of data to evaluate social validity. The legitimacy of several design decisions must be considered. First, although the benchmarks were created by the researcher and not children with disabilities, the credibility of those benchmarks is supported by their grounding in critical literature that draws on children's voices and lived experiences.

Second, the benchmarking process enabled the researcher to evaluate social validity in an unconventional manner; that is, explicitly solicited subjective perspectives regarding the COSA were integrated with data that did not explicitly seek end user's subjective evaluations of the COSA. For instance, Rasch measurement data and open-ended COSA responses were interpreted in alternative ways to evaluate social validity. The threat to the credibility of these interpretations arises from the researcher's role in making inferences rather than making interpretations only from end user's perspectives. However, the COSA data were directly derived from children's responses and the data were juxtaposed against QUAL data that explicitly sought children's subjective evaluations. Therefore, the researcher-driven inferences maintain some credibility as they remained grounded in the children's perspective. In future research, the legitimacy of inferences could be enhanced if researchers and end users collaborate to establish the benchmarks used to evaluate self-report social validity.

Although equal weight was placed on each piece of QUAL and QUANT evidence used to evaluate social validity benchmarks (Creswell et al., 2003), QUAL evidence was used *more often* than QUANT evidence. Therefore, results stemming from the QUAL study had a greater influence on the inferences made regarding the social validity of the COSA. The QUAL phase did use numerous strategies to triangulate data within and across children to enhance the trustworthiness and representativeness of the QUAL findings. Consequently, the legitimacy of inferences regarding social validity may be enhanced by the rigor of the QUAL phase. Furthermore, the constructivist epistemology underlying QUAL methods resonates with the underlying purpose of the COSA to obtain children's perspectives. Therefore, the increased emphasis on QUAL data may be appropriate when evaluating the social validity of self-reports such as the COSA. Other

assessment developers may consider asking the following question when selecting among multiple methods: "Does this method or type of data resonate with the underlying purpose of this self-report assessment?" This question can further help assessment developers optimize the design decisions made given time, funding, and recruitment restrictions of their study (Bryman, 2007).

The third legitimacy issue is the manner in which data were integrated, and in some instances, converted, in the data displays. The data displays created for this study took advantage of the inductive/deductive nature of the original data by integrating data theoretically amenable to inductive/deductive shifts (Morgan, 2007). For example, the integration of Rasch item hierarchies with individual respondent rankings capitalized on the Rasch assumption that the item construct is a continuum (Smith et al., 2002); this continuum lends itself to qualitative comparisons with individually ranked COSA responses. Similarity, the qualitative data selected for *quantitizing* were not extended narratives describing complex emotions, but short answers to focused, open-ended questions on the COSA that could be grouped into content areas for which frequencies could be generated. The assumptions underlying these types of data did not inherently conflict with conversion or integration, and resulted in useful integration of QUAL and QUANT results. To successfully use multiple methods during assessment validation, rather than converting any type of data, assessment developers should ensure that appropriate inferences can be made from converted data (Jang, McDougall, Pollon, Herbert, & Russell, 2008; Onwuegbuzie & Johnson, 2006).

## Conclusion

This article illustrated how multiple forms of data, mainly Rasch measurement data and qualitative data, can be juxtaposed and integrated to evaluate an aspect of validity not typically considered in the process of assessment development: social validity. Social validity was evaluated by systematically identifying data that would enable the researcher to endorse a benchmark statement associated with each facet of social validity: procedures, goals, and outcomes. This process juxtaposed and integrated results stemming from the analysis of rating scales, open-ended responses, behavioral observations, and interviews. Data were mixed at the level of inference or integrated using data displays and mixed at the analytical level.

Assessment developers interested in using multiple methods to establish social validity should carefully consider the decisions made regarding sampling strategies, the influence of different forms of data on overall inferences, and the conversion and integration of multiple forms of data. These decisions can affect the legitimacy of the evaluation of social validity. Legitimacy may be enhanced by considering the extent to which multiple samples are appropriate given the research question, reflect the assessment's target population, and are hypothesized to vary in their responses to the assessment. Legitimacy may also be enhanced by using data collection methods that are aligned with the underlying purpose of the assessment.

### Author's Note

This study was conducted at the University of Illinois at Chicago.

### Acknowledgment

The author wishes to thank the children who took the time to share their stories, and all the researchers and clinicians who used the COSA and shared responses for analysis in this study. The author also expresses her thanks Sharon Foster, PhD, Sara J. McLaughlin, PhD, and Brenda H. Vrkljan, PhD, for their thoughtful feedback on the earlier versions of this article.

## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

## Funding

This study was funded with the support of the American Occupational Therapy Foundation, the University of Illinois at Chicago Graduate College, and the National Institute of Disability and Rehabilitation Research ARRT (Grant No. H133P050001).

## Notes

1. Examples of modifications include reading an item for a child, circle the child's response if he or she is physically unable, or providing additional, individualized examples to facilitate understanding of items.
2. In the results of this study, this group is referred to as Competence Class 1 and Importance Class 1.
3. In the results of this study, this group is referred to as Competence Class 2 and Importance Class 2.

## References

American Occupational Therapy Association. (2002). Occupational therapy practice framework: Domain and process. *American Journal of Occupational Therapy, 56,* 609-639.

Betz, C. L., Redcay, G., & Tan, S. (2003). Self-reported health care self-care needs of transition age youth: A pilot study. *Issues in Contemporary Pediatric Nursing, 26,* 159-181.

Bryman, A. (2007). Barriers to integrating quantitative and qualitative research. *Journal of Mixed Methods Research, 1*(1), 8-22.

Cavet, J., & Sloper, P. (2004). Participation of disabled children in individual decisions about their lives and in public decisions about service development. *Children & Society, 18,* 278-290.

Creswell, J. W., Plano Clark, V. L., Gutmann, M. L., & Hanson, W. E. (2003). Advanced mixed methods research designs. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social and behavioral research* (pp. 209-240). Thousand Oaks, CA: SAGE.

Dellinger, A. B., & Leech, N. L. (2007). Toward a unified validation framework in mixed methods research. *Journal of Mixed Methods Research, 1*(4), 309-332.

Fasoli, L. (2003). Reading photographs of young children: looking at practices. *Contemporary Issues in Early Childhood, 4,* 32-47.

Fawcett, S. B. (1991). Social validity: A note on methodology. *Journal of Applied Behavior Analysis, 24,* 235-239.

Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research issues and procedures. *Journal of Consulting and Clinical Psychology, 67,* 308-319.

Gan, C., Campbell, K. A., Snider, A., Cohen, S., & Hubbard, J. (2008). Giving Youth a Voice (GYV): A measure of youth's perceptions of the client-centredness of rehabilitation services. *Canadian Journal of Occupational Therapy, 75,* 96-104.

Greene, J. (2008). Is mixed methods social inquiry a distinctive methodology? *Journal of Mixed Methods Research, 2*(1), 7-22.

Hemmingsson, H., Kottorp, A., & Bernspång, B. (2004). Validity of the school setting interview: an assessment of the student-environment fit. *Scandinavian Journal of Occupational Therapy, 11,* 171-178.

Hitchcock, J. H., Nastasi, B. K., Dai, D. Y., Newman, J., Jayasena, A., Bernstein-Moore, R., . . . Varjas, K. (2005). Illustrating a mixed-method approach for validating culturally specific constructs. *Journal of School Psychology, 43,* 259-278.

Jang, E. E., McDougall, D. E., Pollon, D., Herbert, M., & Russell, P. (2008). Integrative mixed methods data analytic strategies in research on school success in challenging circumstances. *Journal of Mixed Methods Research, 2*(3), 221-247.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Keller, J., Kafkes, A., Basu, S., Federico, J., & Kielhofner, G. (2005). *The Child Occupational Self Assessment (COSA)* Version 2.1. Chicago, IL: MOHO Clearinghouse, Department of Occupational Therapy, College of Applied Health Sciences, University of Illinois at Chicago.

King, G., Law, M., King, S., Hanna, S., Kertoy, M., Rosenbaum, P., & Young, N. (2004). *Children's Assessment of Participation and Enjoyment (CAPE) and Preferences for Activities of Children (PAC)*. San Antonio, TX: Harcourt Assessment.

Kramer, J., & Hammel, J. (IN PRESS). "I do lots of things:" Children with cerebral palsy's competence for everyday activities. *International Journal of Disability, Development, and Education*.

Kramer, J., Smith, E. V., Jr., & Kielhofner, G. (2009). Rating scale use by children with disabilities on a self report of everyday activities. *Archives of Physical Medicine and Rehabilitation, 90,* 2047-2053.

Lee, Y. J., & Greene, J. (2007). The predictive validity of an ESL placement test: A mixed methods approach. *Journal of Mixed Methods Research, 1*(4), 366-389.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47,* 149-174.

McConney, A., Rudd, A., & Ayres, R. (2002). Getting to the bottom line: A method for synthesizing findings within mixed-methods program evaluations. *American Journal of Evaluation, 23,* 121-140.

Messick, S. (1995). Validity of Psychological Assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741-749.

Messick, S. (1998). Test validity: A matter of consequence. *Social Indicators Research, 45,* 35-44.

Miles, M. B., & Huberman, A. B. (Eds.). (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). Thousand Oaks, CA: SAGE.

Morgan, D. L. (2007). Paradigms lost and pragmatism regained: methodological implications of combining qualitative and quantitative methods. *Journal of Mixed Methods Research, 1*(1), 48-76.

Morse, J. M. (1998). Designing funded qualitative research. In N. K. Denzin & Y. S. Lincoln (Eds.), *Collecting and interpreting qualitative materials* (pp. 56-85). Thousand Oaks, CA: SAGE.

Myford, C. M., & Mislevy, R. J. (1995). *Monitoring and improving a portfolio assessment system*. Princeton, NJ: Center for Performance Assessment.

Onwuegbuzie, A. J., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *The Qualitative Report, 12,* 281-316.

Onwuegbuzie, A. J., & Johnson, R. B. (2006). *The validity issue in mixed research. Research in the Schools, 13,* 48-63.

Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: SAGE.

Rebok, G., Riley, A., Forrest, C., Starfield, B., Green, B., Robertson, J., & Tambor, E. (2001). Elementary school-aged children's reports of their health: A cognitive interviewing study. *Quality of Life Research, 10*, 59-70.

Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement, 14,* 271-282.

Smith, E. V., Jr., Conrad, K. M., Chang, K., & Piazza, J. (2002). An introduction to Rasch measurement for scale development and person assessment. *Journal of Nursing Measurement, 10,* 189-206.

Strauss, A., & Corbin, J. (1998). Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Strategies of qualitative inquiry* (pp. 158-183). Thousand Oaks, CA: SAGE.

Sturgess, J., Rodger, S., & Ozanne, A. (2002). A review of the use of self-report assessment with young children. *British Journal of Occupational Therapy, 65,* 108-116.

Taylor, S. J., & Bogdan, R. (1998). *Introduction to qualitative research methods: A guidebook and resource* (3rd ed.). New York, NY: Wiley.

Teddlie, C., & Yu, F. (2007). Mixed methods sampling: A typology with examples. *Journal of Mixed Methods Research, 1*(1), 77-100.

Turner, C. (2003). *Are you listening? What disabled children and young people in Wales think about the services they use*. Cardiff, Wales: Barnardos, Children First, NCH.

Valnes, B., Hare, R., Block, J., Branch, C., & Broshears, G. (2003). *Future directions agenda for youth with disabilities, 2003*. Portland, OR: National Youth Leadership Network.

Wagner-Menghin, M. M. (2006). The Mixed-Rasch Model: An example for analyzing the meaning of response latencies in a personality questionnaire. *Journal of Applied Measurement, 7*(2), 225-237.

Wolf, M. M. (1978). Social validity: The case for subjective measurement, or, how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11,* 203-214.

World Health Organization. (2001). *International classification of functioning, disability, and health*. Geneva, Switzerland: Author.

Wright, B. D., & Linacre, M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 70,* 857-860.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Young, B., Rice, H., Dixon-Woods, M., Colver, A. F., & Parkinson, K. N. (2007). A qualitative study of the health-related quality of life of disabled children. *Developmental Medicine & Child Neurology, 49,* 660-665.