

---

# Validity Evidence for the Child Occupational Self Assessment

Jessica M. Kramer, Gary Kielhofner, Everett V. Smith, Jr.

## KEY WORDS

- activities of daily living
- disabled children
- psychometrics
- self assessment (psychology)

To ensure that researchers and practitioners make valid interpretations from children's self-reports, evidence must demonstrate that the self-report has appropriate psychometric properties. The Child Occupational Self Assessment (COSA) is a self-report of occupational competence and value for everyday activities designed to involve children in identifying goals and assessing outcomes. Five hundred two children with disabilities, ages 6–17, completed the COSA. We used a Rasch Partial Credit model and parametric and non-parametric statistics to obtain validity evidence. Evidence suggests that the COSA has good content, structural, and substantive validity as given by item and child fit statistics and unidimensionality evaluation. Evidence for external validity was mixed because child fit status and measures varied with some demographic and assessment administration variables. Evidence suggests that most children's responses to the COSA can be validly interpreted as indicators of occupational competence and value for everyday activities.

Kramer, J. M., Kielhofner, G., & Smith, E. V., Jr. (2010). Validity evidence for the Child Occupational Self Assessment. *American Journal of Occupational Therapy, 64*, 621–632. doi: 10.5014/ajot.2010.08142

**Jessica M. Kramer, PhD, OTR/L**, is Assistant Professor, Department of Occupational Therapy, Boston University, 635 Commonwealth Avenue, SAR 512, Boston, MA 02215; kramerj@bu.edu

**Gary Kielhofner, DrPH, OTR/L, FAOTA**, is Professor, Wade/Meyer Chair, Department of Occupational Therapy, University of Illinois at Chicago.

**Everett V. Smith, Jr., PhD**, is Associate Professor, Department of Educational Psychology, University of Illinois at Chicago.

The field of occupational therapy has developed several self-report assessments for children and youths, including assessments of motor control competence (Missiuna, Pollock, & Law, 2004), participation in everyday activities (Keller, Kafkes, Basu, Federico, & Kielhofner, 2005; Law et al., 1998), and leisure activities (Henry, 2000; King et al., 2004). Researchers and practitioners can use these assessments to identify client concerns, develop meaningful goals for intervention, and generate client-reported outcomes. However, to use children's self-reports in practice with confidence, evidence must demonstrate that self-reports actually assess the concepts they propose to measure and that measurement of those concepts is consistent across practice contexts, administration methods, and children with different needs and abilities. This evidence supports the validity of interpretations made by practitioners and researchers on the basis of self-report responses.

The Child Occupational Self Assessment (COSA; Keller, Kafkes, Basu, et al., 2005) is a self-report of occupational competence and value for everyday activities. The COSA is based on concepts from the Model of Human Occupation (MOHO; Kielhofner, 2008) and measures the extent to which children feel they competently meet expectations and responsibilities associated with activities and the relative value of those activities. The process of completing the COSA also facilitates children's involvement in the therapy process. The administration manual directs practitioners to attend to those activities that the child feels less competent doing but for which he or she indicates high importance; those activities are priorities that should be addressed in therapy. To ensure the validity of interpretations, evidence must demonstrate that the COSA items are appropriate measures of the concepts of occupational competence and value and that children's interpretations of the items representing those constructs do not vary across administration contexts or personal demographics such as disability or gender.

Previous research suggested that some COSA items may not be accurate measures of occupational competence and values (Keller, Kafkes, & Kielhofner, 2005; Keller & Kielhofner, 2005). In addition, children did not interpret two COSA items in the intended manner (“Use money to buy things by myself” and “Move my body from one place to another”), and those items were revised. In this study, we analyze a revised version of the COSA using a larger sample to build further validity evidence and determine whether revisions resolved previous item problems.

Several concepts underlie the broad concept of validity (Messick, 1995), and in combination, evidence for each of these concepts supports the COSA’s valid use and interpretation. *Content validity* addresses the representativeness of the measured concepts and the technical quality of the items representing those concepts (Wolfe & Smith, 2007). Evidence of content validity signifies that the items presented on the COSA are easy for children to understand and interpret and appropriately represent the range of everyday activities that contribute to a child’s sense of competence and value.

*Structural validity* refers to the relationship between the assessment outcome and the theory underlying the concept being assessed (Wolfe & Smith, 2007). The COSA intends to measure two concepts based on MOHO: occupational competence and value. Therefore, evidence of structural validity would support the interpretation that a child’s combined response to one rating scale is an indication of his or her underlying sense of competence or value for everyday activities.

*Substantive validity* is the extent to which theory explains differences in response to items (Wolfe & Smith, 2007). This aspect of validity is crucial because practitioners use theory to interpret a child’s responses to the COSA. For example, children who report moderate levels of competence for everyday activities that are easier to perform should report lower levels of competence for everyday activities that are more difficult to perform. These children are also assumed to have less competence than children who consistently report high competence for everyday activities. This theoretical interpretation is supported by evidence of substantive validity; this validity ensures that practitioners can interpret responses in a meaningful way to make appropriate intervention plans and address children with disabilities’ competence and value for everyday activities.

*External validity* is typically established by comparing responses to one assessment with responses to another assessment. Evidence for external validity, however, can also be gathered by examining group differences on external variables (Wolfe & Smith, 2007). For example, the COSA

was designed to be used with children and youths ages 7–17 with a wide range of abilities and needs. Therefore, evidence should suggest that children and youth representing a range of ages, diagnoses, and practice contexts (e.g., school, community-based clinic) should interpret and respond to the COSA in a consistent manner. Moreover, to ensure that children can access and complete the self-report process, the COSA includes several administration options. The use of those administration methods should not lead to differences in children’s understanding of the assessment items or rating scales. Finally, differences in reported levels of competence and value for everyday activities should not be related to general group variable, such as age or diagnosis. Rather, reported levels of competence and value are expected to be related to personal and environmental factors that may be more likely to facilitate or restrict children’s choice of and active engagement in activities. Evidence of external validity helps practitioners make valid interpretations of COSA responses with different client groups and across practice contexts.

The purpose of this study was to provide evidence for each aspect of validity we have described.

## Method

### *Child Occupational Self Assessment*

The COSA consists of a series of 25 items that represent a range of everyday activities that most children encounter at home, at school, and in their communities. Each item is rated using two rating scales: the Occupational Competence scale and the Values scale (Figure 1).

The COSA can be administered in one of three ways, depending on each child’s abilities: (1) a standard paper-and-pencil form that provides different visual cues (faces and stars) for each rating category description (see Figure 1), (2) a card-sort version that places each item on a separate card and each rating category (and visual cue) on a larger rating card, and (3) a summary form that presents all items and rating categories in a matrix format without visual cues. The summary form was originally developed for practitioners to use when recording and reporting children’s responses, but previous research indicated that some children preferred to use the form without visual cues (Keller & Kielhofner, 2005). During administration, all children are reminded to pick the category that best describes how they feel and that there is no right or wrong answer.

COSA items were designed to provide children with a structure for thinking about a range of everyday activities while also allowing children to insert individualized concerns and needs into the assessment process. For

Myself	I have a big problem doing this	I have a little problem doing this	I do this ok	I am really good at doing this	Not really important to me	Important to me	Really important to me	Most important of all to me
Get enough sleep	☹ ☹	☹	☺	☺ ☺	★	★★	★★★	★★★★
Have enough time to do things I like	☹ ☹	☹	☺	☺ ☺	★	★★	★★★	★★★★
Take care of my things	☹ ☹	☹	☺	☺ ☺	★	★★	★★★	★★★★
Get around from one place to another	☹ ☹	☹	☺	☺ ☺	★	★★	★★★	★★★★
Choose things that I want to do	☹ ☹	☹	☺	☺ ☺	★	★★	★★★	★★★★

**Figure 1. Excerpt from the Child Occupational Self Assessment.**

Note. From *The Child Occupational Self Assessment (COSA) (Version 2.1)* (p. 3), by J. Keller, A. Kafkes, S. Basu, J. Federico, and G. Kielhofner, 2005, Chicago: University of Illinois, College of Applied Health Sciences, Department of Occupational Therapy, MOHO Clearinghouse. Copyright © 2005 by the MOHO Clearinghouse. Used with permission.

example, on the item “Get my chores done,” one child may think about feeding his dog, and another child may think about cleaning her room. This flexibility enables practitioners to identify goals and intervention activities that are unique to each child. However, evidence must demonstrate that even with this flexibility, items on the COSA consistently measure one underlying concept.

### Rasch Partial Credit Model Analysis

We used the Rasch Partial Credit (PC) model (Masters, 1982; Wright & Masters, 1982) to estimate the extent to which each item represents the underlying concept, or construct, it intends to measure and the extent to which each child possessed each underlying construct measured. These estimates are given in logits and are interval units of measurement when the data fit model expectations.

When items meet the criteria for the PC model, or “fit,” they are determined to be valid representations of the underlying construct. Similarly, when children’s responses fit model expectations, their responses are internally consistent and can be interpreted as valid measures of the underlying constructs. A mean square fit statistic (*MnSq*) can be used to assess the fit of items and children to model expectations; each *MnSq* is also accompanied by a stan-

dardized mean square (*Zstd*), which indicates statistical significance (Hashway, 1978; Wright & Stone, 1979). The ideal value for the *MnSq* is 1.0. *MnSq* values >1.0 indicate more variability than expected by the model. In this study, *MnSq* >1.4 that are also associated with *Zstd* ≥ 2.0 indicate poor fit as recommended by Wright and Linacre (1994) and in keeping with the development of other occupational therapy assessments (Bray, Fisher, & Duran, 2001; Velozo et al., 1999).

The PC model also tests whether the items of each scale coalesce to form a single dominant construct, the assumption of unidimensionality (Hashway, 1978; Wright & Stone, 1979). The COSA has two rating scales and two hypothesized unidimensional constructs: occupational competence and value for everyday occupations. A three-step process outlined by Smith (2002, 2004) acts as a rigorous test for departures from unidimensionality. First, the PC analysis, as implemented in the Winsteps Version 3.49 software program (Linacre, 2004), provides a point-biserial correlation coefficient for each item. Negative item point-biserial correlation coefficients may indicate multidimensionality (Smith, 2002). Second, poor item fit, as described previously, may signify multidimensionality. Third, a principal components analysis

(PCA) of the item residuals will be investigated for thematic patterns of items that may indicate multidimensionality. Additional information about dimensionality is gained by examining the amount of unexplained residual variance found in the PCA. When items are multidimensional, the residual variance represents more than random noise. To determine whether the amount of residual variance represents more than random noise, we compare the PCA of the COSA dataset with a PCA of simulated data that are known to be unidimensional. The simulated data provide an estimate of the amount of residual variance expected when the COSA data are unidimensional and perfectly fit model expectations; if the residual variance explained by the first component of the COSA dataset is substantially greater than this estimate, data may be multidimensional.

The PC model also yields item measures that identify the location of each item along the continuum, or hierarchy, representing the underlying construct. On the COSA, items with higher measures are more difficult for children to do well or are less important, and items with smaller measures are easier for children to perform or are more important. Within the PC model, one can use a technique called *pivot anchoring* to compare the overall item measures in a meaningful way. Pivot anchoring allows the researcher to set the definition of each item's difficulty around a theoretically driven "pivot" point in the rating scale. For the COSA, the pivot for all Occupational Competence items was the transition between responses indicating less competence ("I have a little problem doing this") and more competence ("I do this ok"). For all Values item, the pivot was the transition between responses that indicate no importance ("Not really important to me") and importance ("Important to me"). This technique results in a more interpretable item hierarchy. This item hierarchy is an indication of how most children view the difficulty and importance of everyday activities. To help establish validity, the distribution of items across the hierarchy should be logical. For example, we would expect children to report that the item "Do things with my family" is easier to do than the item "Calm myself down when I am upset" (Keller & Kielhofner, 2005).

Winsteps also provides an estimate of the number of statistically different levels of occupational competence and value represented by the continuum of items; this is known as the *number of strata* (Wright & Masters, 1982). Strata are derived using the formula

$$(4[\text{separation index}] + 1)/3.$$

More strata indicate that the items separate to measure a broader range of the underlying construct. The reliability

of that separation is also provided by Winsteps; item separation reliability is the ratio of adjusted variance to observed variance and represents the proportion of variance not resulting from error (Smith, 2001).

Similarly to items, the PC model also estimates the position of each child on a continuum from less to more of the construct. Children with higher measures are those assumed to have a greater sense of occupational competence or value for everyday activities. The PC model is robust to missing data and is able to estimate a child's measure even when some items are missing. Because children are not required to answer every COSA item, the use of this model ensures that the analysis includes as many children's responses as possible. Winsteps also provides a person separation reliability that represents the proportion of variance not resulting from error and is analogous to Cronbach's  $\alpha$  (Smith, 2001).

In this study, we asked the following research questions to build the validity evidence for the COSA:

1. Does the COSA have evidence of content validity, as given by acceptable item fit statistics?
2. Does the COSA have evidence of structural validity, as given by evidence of unidimensionality?
3. Does the COSA have evidence of substantive validity, as given by acceptable child fit statistics and theoretically congruent item hierarchies?
4. Does the COSA have evidence of external validity, as given by a lack of group differences in child fit and measures across administrative variables and personal demographics?

### *Procedure*

A university research ethics committee approved this research protocol. Researchers and clinicians from around the world were contacted through e-mail distribution lists and Web site announcements and during professional meetings and training and invited to share children's COSA responses with a central database. Ninety-eight practitioners and researchers who were occupational therapists and physical therapists contributed COSA data. The COSA has been translated into several languages by practitioner-researchers for use in practice. Icelandic, German, and Italian translations were included in this study. Details about the process used to translate each assessment are available in the online supplemental materials for this article (go to [www.ajot.ajotpress.net](http://www.ajot.ajotpress.net), navigate to this article, and click on "supplemental materials").

Most COSAs were completed as part of standard practice (85.2%), and the remainder were collected as part of other research studies. These researchers and clinicians administered the COSA by following the directions in the

administration manual (Keller, Kafkes, Basu, et al., 2005) and made individualized administrative modifications as needed. Children completed the COSA in an average of 27 min (range = 5–135 min). COSA responses and basic deidentified demographic and administration information were provided to the central database; because of the nature of data collection, some demographic information was missing or unavailable even after follow-up with the original contributor. Not all children completed both scales as a result of fatigue, lack of time, or administrative modifications.

The COSA rating scale responses were converted to the numbers 1–4 for data entry, with 1 assigned to the lowest category for each rating scale. The data from the COSA’s two scales were analyzed using the Rasch PC model as implemented by Winsteps (Linacre, 2004). A PC model analysis without pivot anchoring was run first to derive information about children. Group differences were then explored in several ways using SPSS Version 12.0 (SPSS, Inc., Chicago). We used  $\chi^2$  tests to examine the relationship between children’s fit status (designated as 1 = *good fit* or 2 = *poor fit*) and demographic or administration context variables. The relationship between children’s measures, demographics, and administration context was explored using *t* tests and one-way analysis of variance (ANOVA); Bonferroni corrections were made when appropriate. Pivot anchoring was then used during a second PC model analysis to obtain meaningful item hierarchies (Bode, 2004).

### Sample

Five hundred two children responded to the Occupational Competence scale, and 496 of those children also responded to the Values scale (Table 1). The most common developmental delay was a specific delay in development such as expressive language delay ( $n = 119$ ), followed by autism ( $n = 74$ ) and attention deficit disorder ( $n = 68$ ). The most common neurological diagnosis was cerebral palsy ( $n = 27$ ), followed by spina bifida ( $n = 11$ ). In addition to their primary diagnosis, 100 children had intellectual disabilities.

Most children completed the COSA using the standard form with faces and stars ( $n = 361$ ), but some used the summary form ( $n = 75$ ) or the card sort ( $n = 63$ ). Practitioners provided modifications during the assessment administration for 162 children; examples of modifications included using a cover sheet to reduce visual distractions, providing individualized examples for items when a child did not understand, and writing on behalf of the child when he or she had difficulty circling the desired response category. The database included responses from Icelandic (6.8%), German (4.4%), and Italian (1.2%) translations of the COSA. Some

**Table 1. Children’s Demographic Information**

Demographics	<i>n</i>	%
Gender		
Male	360	71.7
Female	142	28.3
Ethnicity <sup>a</sup>		
White	410	81.7
African-American/African	35	7.0
Hispanic	22	4.4
Asian/Pacific Islander	10	2.0
Middle Eastern	10	2.0
Multiracial	10	2.0
Other	4	0.8
Missing information	1	0.2
Major condition or diagnostic category <sup>a</sup>		
Developmental delay	323	64.3
Neurological	88	17.5
Mental health	46	9.2
Unknown or other diagnosis	15	3.0
No diagnosis, but receiving occupational therapy	12	2.4
Orthopedic/musculoskeletal	10	2.0
Medical/chronic health	6	1.2
Visual/hearing/sensory impairment	1	0.2
Missing information	1	0.2
Practice setting		
School (inclusive setting)	233	46.4
School (institutional/center setting)	144	28.7
Outpatient/community-based rehabilitation	92	18.3
Inpatient hospital	24	4.8
Other setting	9	1.8
Country <sup>a</sup>		
United States	253	50.4
United Kingdom	187	37.3
Iceland	34	6.8
Germany	22	4.4
Switzerland	6	1.2
Age ( $n = 497$ )		
Mean = 11 yr 11.73 mo (standard deviation = 2 yr 10.39 mo)		
Range = 6 yr–17 yr 10 mo		

Note.  $N = 502$ .

<sup>a</sup>Does not add up to 100; percentages rounded to one decimal place for interpretability.

children used the original English version supplemented with a British Sign Language translation (1.4%).

## Results

### Items

We examined item fit, unidimensionality, and the relative ordering of items along the competence and value item hierarchies to answer the first three research questions.

**Occupational Competence Items.** All Occupational Competence items met the Rasch model fit requirements (Table 2). All Occupational Competence items had positive point-biserial correlations ranging from .30 to .53. The PCA of standardized residuals grouped together

**Table 2. Occupational Competence Item Measures and Fit Statistics**

Occupational Competence Item <sup>a</sup>	Measure	Error	Infit		Outfit	
			<i>MnSq</i>	<i>Zstd</i>	<i>MnSq</i>	<i>Zstd</i>
Keep my mind on what I am doing	0.70	.05	1.02	0.30	1.04	0.60
Get enough sleep	0.60	.05	1.11	2.00	1.14	2.00
Calm myself down when I am upset	0.48	.05	1.00	0.10	0.99	-0.10
Think of ways to do things when I have a problem	0.47	.05	0.97	-0.50	0.96	-0.60
Get my homework done	0.43	.05	1.04	0.60	1.08	1.10
Keep working on something even when it gets hard	0.35	.05	0.94	-1.00	0.99	-0.10
Get my chores done	0.35	.05	1.01	0.20	0.99	-0.10
Make others understand my ideas	0.28	.05	0.93	-1.20	0.93	-1.00
Finish what I am doing without getting tired too soon	0.22	.05	0.92	-1.40	0.94	-0.80
Finish my work in class on time	0.21	.06	0.92	-1.50	0.92	-1.30
Follow classroom rules	0.16	.05	1.05	0.80	1.08	1.00
Buy something myself	0.15	.05	1.10	1.70	1.20	2.40
Make my body do what I want it to do	0.14	.05	0.92	-1.20	0.94	-0.60
Ask my teacher questions when I need to	0.04	.06	1.03	0.50	0.98	-0.20
Take care of my things	-0.09	.06	0.97	-0.50	0.95	-0.50
Do things with my classmates	-0.23	.06	0.93	-1.10	0.93	-0.90
Have enough time to do things I like	-0.26	.05	1.07	1.10	1.27	3.10
Choose things that I want to do	-0.34	.06	0.95	-0.70	0.98	-0.20
Use my hands to work with things	-0.35	.06	1.03	0.40	1.01	0.20
Keep my body clean	-0.43	.06	0.98	-0.30	1.11	1.20
Get around from one place to another	-0.49	.06	1.03	0.40	1.12	1.20
Eat my meals without any help	-0.50	.07	1.10	1.10	1.16	1.20
Do things with my family	-0.52	.06	1.04	0.60	1.14	1.50
Dress myself	-0.63	.06	1.06	0.80	1.11	1.10
Do things with my friends	-0.71	.06	0.96	-0.50	1.03	0.40

Note. *MnSq* = mean square fit statistic; *Zstd* = standardized mean square.

<sup>a</sup>Items with higher measures are more difficult for children to indicate high competence on. Items with lower measures are easier for children to indicate high competence on.

most of the “easier” (lower measures) and “more difficult” (higher measures) Occupational Competence items. In the COSA dataset, 4.2% of the residual variance was explained by the first component, compared with 2.6% explained by the first component of the simulated dataset.

The Occupational Competence item hierarchy shows that basic activities of daily living (ADLs) and spending time with family and friends are the activities for which most children were likely to indicate high competence. Performing classroom-related tasks and managing one’s body, personal things, and schedule were tasks most children reported moderate competence for doing relative to other items. Children reported the least amount of competence and the most difficulty for self-regulation and cognitive tasks, as well as chores. Across the continuum of Occupational Competence, the item separation index was 6.18, which translates to 8.57 strata; the reliability of item separation was .97.

*Values Items.* Twenty-four of the 25 Values items met the fit requirements; the item “Get my chores done” did not demonstrate goodness of fit (Table 3). All Values item had positive point-biserial correlations ranging from .42 to .61. Similar to the Occupational Competence

scale, the PCA of standardized residuals grouped together most of those Values items that children reported to be most important and the Values items that children reported to be least important. The PCA of standardized residuals of the COSA dataset resulted in 4.4% of residual variance explained by the first component, more than the 1.8% explained by the first component of the simulated data.

The Values item hierarchy depicts a continuum that ranges from personal tasks to tasks required by others. Children were most likely to indicate the highest value for tasks and activities they personally controlled, such as “Take care of my things” and “Make my body do what I want it to do.” Children were less likely to indicate value for activities typically regulated and demanded by adults such as “Get my chores done” and “Follow classroom rules.” Values item separation across the continuum was 3.96, which transforms to 5.6 strata; item separation reliability across that continuum was 0.94.

### Children

We examined child fit and the relationship between the COSA responses and demographic or administration

**Table 3. Values Item Measures and Fit Statistics**

Values Item <sup>a</sup>	Measure	Error	Infit		Outfit	
			<i>MnSq</i>	<i>Zstd</i>	<i>MnSq</i>	<i>Zstd</i>
Get my chores done <sup>b</sup>	1.07	.05	1.22	3.60	1.40	4.90
Get my homework done	0.72	.05	1.00	0.00	1.00	0.10
Keep my body clean	0.32	.05	1.11	1.90	1.22	2.90
Follow classroom rules	0.31	.05	1.04	0.70	0.99	-0.20
Buy something myself	0.29	.05	1.20	3.40	1.34	4.50
Eat my meals without any help	0.29	.05	1.08	1.30	1.18	2.10
Do things with my classmates	0.16	.05	0.96	-0.70	0.95	-0.70
Do things with my family	0.06	.05	1.13	2.10	1.18	2.10
Get around from one place to another	0.03	.05	0.97	-0.40	0.97	-0.50
Make others understand my ideas	-0.01	.05	0.90	-1.70	0.87	-2.10
Get enough sleep	-0.02	.05	1.08	1.30	1.34	4.00
Ask my teacher questions when I need to	-0.02	.05	0.88	-2.20	0.89	-1.70
Do things with my friends	-0.07	.05	1.13	2.10	1.20	2.40
Keep working on something even when it gets hard	-0.09	.05	0.91	-1.70	0.98	-0.40
Finish what I am doing without getting tired too soon	-0.10	.05	0.86	-2.70	0.81	-3.10
Calm myself down when I am upset	-0.11	.05	1.02	0.40	1.15	2.10
Finish my work in class on time	-0.12	.05	0.87	-2.40	0.84	-2.60
Choose things that I want to do	-0.15	.05	1.02	0.40	1.01	0.20
Keep my mind on what I am doing	-0.18	.05	0.87	-2.40	0.87	-2.00
Think of ways to do things when I have a problem	-0.19	.05	0.91	-1.60	0.95	-0.70
Dress myself	-0.27	.05	1.03	0.50	1.04	0.50
Use my hands to work with things	-0.32	.05	0.86	-2.60	0.80	-2.90
Make my body do what I want it to do	-0.36	.06	0.93	-1.10	0.90	-1.30
Have enough time to do things I like	-0.45	.05	1.16	2.70	1.28	3.60
Take care of my things	-0.79	.06	0.95	-0.80	0.92	-1.10

Note. *MnSq* = mean square fit statistic; *Zstd* = standardized mean square.

<sup>a</sup>Items with higher measures are less likely to be perceived as important. Items with lower measures are more likely to be perceived as important.

<sup>b</sup>Item exceeds fit criteria: *MnSq* ≥ 1.4 associated with *Zstd* ≥ 2.0.

context variables to answer the third and fourth research questions.

**Occupational Competence.** Of the 502 children who completed the Occupational Competence ratings, 59 did not meet fit requirements (11.75%). Seven children responded using only the highest rating category and therefore were not assigned fit statistics; these children were not included in the following analysis. The reliability with which children were measured along the Occupational Competence continuum was .83.

The children who exceeded fit requirements were compared with those who had acceptable fit statistics on a variety of demographics and aspects of assessment administration (Table 4). Fit status was not statistically significantly different by the child's age, gender, ethnicity, diagnostic category, country, practice setting, administration time, or use of a COSA translation. For variables that had <5 observations in cells of the  $\chi^2$  table, categories were combined and variables underwent a second analysis. Only the use of a COSA translation had a significant  $\chi^2$  during this second analysis; children who used any translation of the COSA were more likely to have good fit than children who used the original English ( $\chi^2[1, N =$

495] = 4.4,  $p = .036$ ). Children with poor fit were statistically more likely to use the card sort method, have an intellectual disability, and use modifications during administration. There was also a statistically significant difference between the measures of children who had good versus poor fit, with children with poor fit receiving significantly lower Occupational Competence measures than children with good fit.

Children's Occupational Competence measures were not statistically related to any of the demographic categories or assessment administration contexts. All  $t$  tests exploring the relationship between Occupational Competence measures and gender, presence of intellectual disability, and use of modifications were nonsignificant. All ANOVAs exploring the relationship between Occupational Competence measures and administration method, ethnicity, diagnostic category, country, use of translation, and practice setting were nonsignificant using a Bonferroni-adjusted  $\alpha$  of .01.

**Values.** Of the 496 children who responded to the Values rating scale, 76 did not meet fit requirements (15.3%). In addition, fit statistics were not calculated for 13 children who responded to every Values item using the

**Table 4. Relationship Between Child Fit Status and Other Variables on the Occupational Competence Scale**

Variable	$\chi^2$	df	n	p
Administrative method (1)	15.48	2	492	.000*
Country (2) <sup>a</sup>	5.50	4	495	.240
Diagnostic category (3) <sup>a</sup>	9.40	7	494	.228
Ethnicity (4) <sup>a</sup>	9.90	9	494	.357
Gender	0.001	1	495	.982
Intellectual disability (5)	11.28	1	493	.001*
Modifications (6)	14.40	1	495	.000*
Practice setting (7) <sup>a</sup>	6.05	4	495	.195
Translated Child Occupational Self Assessment (8) <sup>a</sup>	5.20	4	495	.266

  

	Mean (SD)	t	df	p
Age		1.94	490	.052
Good fit	12 yr 2 mo (39.6 mo)			
Poor fit	11 yr 3 mo (33.0 mo)			
Administration time		0.25	482	.799
Good fit	27.33 min (19.5 min)			
Poor fit	26.65 min (14.4 min)			
Occupational Competence measures		2.36	493	.018*
Good fit	.80 (0.82)			
Poor fit	.53 (0.81)			

Note. *df* = degrees of freedom; *SD* = standard deviation. 1 = standard form, card sort, or summary form; 2 = United States, United Kingdom, Iceland, Germany, and Switzerland; 3 = developmental delay, neurological condition, mental health, orthopedic/musculoskeletal, medical/chronic health, sensory impairment, other/unknown, or no diagnosis but receiving occupational therapy; 4 = White (United States), White/European (international), African-American (United States), African (international), Asian/Pacific Islander (United States), Asian/Pacific Islander (international), Hispanic/Latino or Latina (United States), Middle Eastern (United States), multiracial (all), or other (all); 5 = Presence of intellectual disability (yes or no); 6 = use of modifications (yes or no); 7 = inclusive school setting, institutional school setting, outpatient/community rehabilitation, inpatient hospital, or other; 8 = English, Icelandic, German, British Sign Language, or Italian.

<sup>a</sup> $\chi^2$  table included cells with fewer than five observations.

\*Significant at the .01 level.

highest rating category; these children were not included in the following analysis. The reliability with which children were measured along the Values continuum was .85.

Compared with children who had acceptable fit, children with poor fit were not statistically different with respect to their Values measure, gender, ethnicity, diagnostic category, presence of intellectual disability, country, use of a COSA translation, assessment administration method, use of modifications, administration time, or practice setting (Table 5). For variables that had fewer than five observations in cells of the  $\chi^2$  table, categories were combined and a second analysis was conducted. No statistically significant differences were found. Children who had poor fit on the Values scale were statistically younger than children who had good fit. On average, children with poor fit were 2 years younger than children with good fit.

Children's Values measures varied with some assessment contexts. A statistically significant effect was found between the Values measure and assessment administration method ( $F[2, 490] = 4.84, p = .004$ ). Children who used the standard paper-and-pencil form with visuals and the children who used the card sort reported more value than the children who used the summary form. We also found a statistically significant effect by country ( $F[4, 491] = 5.85, p = .00$ ) and COSA translation ( $F[4, 491] = 6.21, p = .00$ ). Children from the United Kingdom reported less value in comparison with children from Iceland and Germany, and children from the United States reported less value in comparison with children from Germany. Children who used the Icelandic and German translations of the COSA had higher Value measures than the children who used the original English version. Other ANOVAs exploring the relationship between Values measures and ethnicity, diagnostic category, and practice setting were not statistically significant after adjusting the significance level to .01. All *t* tests exploring the relationship between Values measures and gender, presence of intellectual disability, and use of modifications were nonsignificant.

## Discussion

In this discussion, we synthesize the results to determine the level of evidence for the content, structural, substantive, and external validity of the COSA.

Results provide evidence for the COSA's content validity. No items exceeded fit requirements on the Occupation Competence scale, and only one item exceeded fit requirements on the Values scale. These findings indicate an improvement in item fit from previous versions of the COSA, in which two and four items had poor fit to the model on the Occupational Competence and Values scales, respectively, suggesting that the current revised items have better technical quality than previous versions (Keller & Kielhofner, 2005). The item "Get my chores done" had poor fit on the Values scale. The positive value of the item measure indicates that relative to other items, most children assigned less importance to chores, making it one of the most difficult items for children to indicate as important. Therefore, the Rasch model expects that those children assigning the highest level of importance to other items would be most likely to assign a high level of importance to chores, and those children assigning moderate levels of importance to other items would be most likely to assign a low to moderate level of importance to chores. However, the item's outfit statistic indicates that some children assigned unexpectedly low importance to chores relative to the importance assigned

**Table 5. Relationship Between Child Fit Status and Other Variables on the Values Scale**

Variable	$\chi^2$	df	n	p
Administrative method (1)	4.45	2	480	.108
Country (2) <sup>a</sup>	2.05	4	483	.727
Diagnostic category (3) <sup>a</sup>	11.04	7	482	.137
Ethnicity (4) <sup>a</sup>	8.15	9	482	.520
Gender	0.02	1	483	.877
Intellectual disability (5)	2.82	1	481	.093
Modifications (6)	2.00	1	483	.158
Practice setting (7) <sup>a</sup>	3.33	4	483	.504
Translated Child Occupational Self Assessment (8) <sup>a</sup>	3.40	4	483	.494

  

	Mean (SD)	t	df	p
Age		4.23	478	.000*
Good fit	12 yrs, 4 mo (3 yrs, 4 mo)			
Poor fit	10 yrs, 8 mo (2 yrs, 7 mo)			
Administration time		-0.11	470	.913
Good fit	27.33 min (19.1 min)			
Poor fit	27.59 min (18.9 min)			
Values measures		-1.66	481	.097
Good fit	0.442 (0.899)			
Poor fit	0.626 (0.815)			

Note. df = degrees of freedom; SD = standard deviation. 1 = standard form, card sort, or summary form; 2 = United States, United Kingdom, Iceland, Germany, and Switzerland; 3 = developmental delay, neurological condition, mental health, orthopedic/musculoskeletal, medical/chronic health, sensory impairment, other/unknown, or no diagnosis but receiving occupational therapy; 4 = White (United States), White/European (international), African-American (United States), African (international), Asian/Pacific Islander (United States), Asian/Pacific Islander (international), Hispanic/Latino or Latina (United States), Middle Eastern (United States), multiracial (all), or other (all); 5 = Presence of intellectual disability (yes or no); 6 = use of modifications (yes or no); 7 = inclusive school setting, institutional school setting, outpatient/community rehabilitation, inpatient hospital, or other; 8 = English, Icelandic, German, British Sign Language, or Italian.

<sup>a</sup> $\chi^2$  table included cells with <5 observations.

\*Significant at the .01 level.

to other items. These unexpected responses caused the poor item fit. Although poor fit can indicate poor technical quality or construct irrelevance, in this instance poor fit most likely reflects children's limited value for this everyday activity.

Results also provide evidence for the structural validity of the two COSA scales. The combined evidence of item fit, positive point-biserial correlation coefficients, and PCA of the item residuals suggest two unidimensional scales. Although the PCA of item residuals grouped most items on the basis of "difficulty" (i.e., items with higher measures grouped together), the groupings do not appear to be associated with specific item content. For example, on the Occupational Competence scale, the PCA grouped more difficult cognitive items with easier items not related to cognitive processes such as "Do things with my family." Moreover, although the residual variance explained by the first component of the COSA dataset was

larger than that for the simulated data, the amount is negligible for both scales and not a strong indication of multidimensionality.

Results also provide evidence for the COSA's substantive validity. The COSA measured most children in a valid manner as indicated by child fit statistics, and child separation reliability was also good. On both scales, >85% of children responded in a consistent manner. Judgments about the effectiveness of measurement often reference a Type I error rate, where no more than 5% of the respondents have poor fit (Smith, 2002). However, because the purpose of the COSA is to facilitate children's involvement in the therapy process and it is not a high-stakes placement assessment, it is necessary only to demonstrate that most children respond in a manner that meets measurement expectations (Kielhofner, Forsyth, Kramer, & Iyenger, 2009).

The theoretically meaningful item hierarchies found in this study provide further evidence for substantive validity. As in previous studies (Keller, Kafkes, & Kielhofner, 2005; Keller & Kielhofner, 2005), children in this study reported the most competence for relationships and basic self-care activities and less competence for items related to managing responsibilities and sustaining effort. In accordance with MOHO's concept of volition (Kielhofner, 2008), children in this study reported more value for those activities over which they had more control; as in previous studies, chores was the least important activity (Keller & Kielhofner, 2005). The revised items measured a broader range of the constructs in comparison with previous versions (Keller & Kielhofner, 2005) because items were reliably separated into more than six and five significantly different levels of the constructs of Occupational Competence and Values, respectively.

Results provide mixed support for the COSA's external validity because the degree of fit and reported competence and values varied with some demographic and administrative context variables. On the Occupational Competence scale, children with unexpected responses (i.e., poor fit) were more likely to have intellectual disabilities. In addition, children who used modifications or the card sort administration were also more likely to have unexpected responses. Several possible explanations for this finding need to be considered. One possible interpretation is that a child's valid use of the Occupational Competence scale is connected to the child's ability and the corresponding level of support required to complete the assessment. The need to use modifications and the card sort could be related to the child's ability because children with motor, process, or communication and interaction skill difficulties may

require this additional support to participate in the self-report process. Similarly, one could assume that children with intellectual disabilities require additional support to complete a self-report assessment. Therefore, it is possible that children who require more support are more likely to give unexpected responses to the COSA.

An alternative explanation for this finding is that the assistance provided by professionals, as when using modifications or an alternative administrative method, changes children's use and interpretation of the Occupational Competence rating scale. Children receiving assistance and those with intellectual disabilities may also be more influenced by the power imbalances inherent in adult-child interactions (Christensen & James, 2000; Rapley, 2004) and therefore more concerned about the adults' opinion of their responses; this concern could lead children to change their response pattern and increase the likelihood of unexpected responses. It is also possible that children who receive assistance to complete a self-report may perceive themselves as less able; this interpretation is further supported by the finding that children with poor fit had significantly lower measures of occupational competence.

However, both of these interpretations regarding the relationships among the degree of fit, modifications, and ability are called into question given the lack of relationship between these variables on the Values scale. If a child's ability or the provision of support influenced the use of the Occupational Competence scale, one would expect to find similar results on the Values scale; however, we found no such relationship. Therefore, the relationship between Occupational Competence and child ability, support, and adult-child interactions during the administration of the COSA requires further investigation.

Degree of fit on the Values scale appears to vary with age; children with unexpected responses (poor fit) were much younger than children with good fit. It is possible that younger children were unable to differentiate between the levels of importance depicted by the top three rating categories. Other research has shown that younger children are more likely to respond using extreme categories and do not use visually based rating scales as effectively as older children (Chambers & Johnston, 2002; Shields, Palermo, Powers, Fernandez, & Smith, 2005). The level of reported values appears to vary with administration method because children who used the summary form had significantly lower value measures than children who used the standard form or card-sort version. It is possible that the star cues provided on the standard form are more desirable to children and artificially inflate children's reported value level. However, taking both of these findings

into consideration, an alternative interpretation is possible. That is, the children who used the summary form were older; therefore, it is possible that they were better able to prioritize the activities that were most important to them as opposed to those that were only moderately important to them. However, it is difficult to determine which interpretation is more appropriate on the basis of the data available in this study. Therefore, the relationship between the reported values, assessment administration method, and age requires further investigation.

Finally, the relationship between the level of reported values, country, and use of a COSA translation requires further exploration. It is possible that the Icelandic and German translations changed the inherent meaning of the Values rating scale and made the categories more desirable to children, artificially inflating their Value measures. The results also point to possible group differences related to translation use, because children who used translations were less likely to have poor fit on the Occupational Competence scale. In this instance, the validity of these translations is called into question. Conversely, differences in reported values may reflect cultural differences; it is possible that the activities depicted in the COSA are more valued in Iceland and Germany. Future research should explore the COSA's cross-cultural relevance by comparing item hierarchies derived from separate analysis of English and other language versions of the COSA and examining differential item functioning.

### *Clinical Implications*

The item hierarchies found in this study have interesting ramifications for therapists. Quite often, occupational therapy focuses on achieving independence in self-care tasks (Brown, Rodger, Brown, & Roever, 2005); however, children in this study felt quite competent in completing these activities. Children did indicate that cognitive tasks such as paying attention, problem solving, and managing emotions were more important to them than daily self-care activities but that they were difficult to perform competently, suggesting that occupational therapists should reconsider the amount of time spent on ADLs during intervention. Occupational therapists may wish to spend more time teaching cognitive strategies and increasing children's sense of efficacy in handling difficult tasks and emotions (Kielhofner, 2008).

On the basis of this study's findings, practitioners should be aware that younger children and children with intellectual disabilities are more likely to have unexpected response patterns to the COSA. The COSA's ultimate purpose, however, is to identify the concerns a child wishes to address in therapy, not to derive a score on

Occupational Competence and Value. Therefore, practitioners are encouraged to provide whatever supports or modifications are necessary to enable children to complete the COSA and identify goals. To limit the impact that modifications and adult support have on COSA responses, practitioners should use evidence-based modifications when possible. For example, research has suggested that rather than placing response categories horizontally, placing them vertically, with the best or highest rating category at the top, facilitates children's understanding of a rating continuum (Beyer & Knapp, 1986). The COSA card sort version of the rating scale uses a horizontal format.

Moreover, practitioners administering the COSA must be critically reflexive (Finlay & Gough, 2003) and cognizant of the impact that professional attitudes, beliefs, and power have on children's responses. Practitioners can begin the self-report process with the assumption that children's reports are valid and that incongruence between children's and adults' reports may simply reflect a difference in the ways in which children and adults construct an understanding of their world (James, Jenks, & Prout, 1998) rather than an inability of a child to provide an accurate self-report. When reviewing responses, practitioners should also determine whether a child rated an item with reference to a specific task or action. For example, a child may indicate high competence for an item that the practitioner believes is challenging for the child (e.g., "Finish my work in class on time"); however, the child may be referencing a task, action, or specific context in which he or she performs more competently (e.g., "I finish projects in art class on time"). Identifying the child's reference point for item ratings further ensures that assessment interpretation is guided by the child's perspective.

### *Limitations*

This study had several limitations. First, this dataset represents a convenience sample; as a result, readers should interpret the relationships among fit, measures, and administration contexts or demographic variables with caution. Although this study used a large sample, some  $\chi^2$  cells still did not have an adequate number of responses. Moreover, findings may not generalize to the population; therefore, replication is needed. Second, the rigor of the processes used to translate the COSA varied across translations. Therefore, findings related to use of a translated assessment should be interpreted with caution. Future research should investigate the psychometric properties of individual COSA translations to ensure the validity of each translation. Third, administration and the application of modifications was not standardized because

many practitioners and researchers contributed data. Future research should systematically recruit children, ensure that administration is standardized, and randomize administration methods. Because the database included only deidentified basic demographic information, we were unable to explore the relationship between COSA responses and additional variables of interest. Future research should gather additional information about the child's functional level, participation in life activities and roles, environment, and volition to better understand how these variables relate to self-reports of competence and value.

## Conclusion

This study used a rigorous measurement approach and a large, heterogeneous clinical sample to build validity evidence for the COSA. The evidence derived from this study suggests that the COSA has good content, structural, and substantive validity, as given by acceptable item and child fit and unidimensionality evaluation. The revised COSA items are theoretically meaningful representations of the constructs of occupational competence and value for everyday activities. Researchers and practitioners should carefully interpret the responses of children with intellectual disabilities and of younger children and be aware that the use of modifications and alternate administration methods may be associated with children's likelihood of responding in an unexpected manner. In conclusion, researchers and practitioners can use MOHO theory to validly interpret most children's responses to the COSA to better understand children's occupational competence and value for everyday activities. ▲

## Acknowledgments

We thank Fairfax County Public Schools, Virginia; Central Manchester and Manchester Children's University Hospital, Manchester, England; Alderwasley/St. John's/Callow Park College, Belper, England; Occupational Therapy Department, Spurwink Services, Portland, Maine; Kennedy Krieger Institute, Baltimore, Maryland; Croydon Primary Care National Health Service (NHS) Trust, Croydon, England; Southwest London and St. George Mental Health NHS Trust, London, England; Children's Centre Worthing Hospital, Worthing, England; Carilion Clinic Children's Hospital, Roanoke, Virginia; University of Illinois Medical Center, Chicago, Illinois; Sara Harney, Pennsylvania; Fran Allen and Helen Smith, United Kingdom; Dr. Jane O'Brien, University of New England; Maike Wolfe, Germany; Hronn Birgisdottir, Kristjana Olafsdottir, and Erla B. Sveinbjornsd, Iceland; and Sabrina Schwarz, Switzerland.

Funding for this project and the preparation of this manuscript received from 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.

## References

- Beyer, J. E., & Knapp, T. R. (1986). Methodological issues in the measurement of children's pain. *Children's Health Care, 14*, 233–241. doi:10.1207/s15326888chc1404\_7
- Bode, R. K. (2004). Partial Credit Model and pivot anchoring. In E. V. Smith, Jr., & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 279–295). Maple Grove, MN: JAM Press.
- Bray, K., Fisher, A. G., & Duran, L. (2001). The validity of adding new tasks to the assessment of motor and process skills. *American Journal of Occupational Therapy, 55*, 409–415.
- Brown, G. T., Rodger, S., Brown, A., & Roever, C. (2005). A comparison of Canadian and Australian paediatric occupational therapists. *Occupational Therapy International, 12*, 137–161. doi:10.1002/oti.2
- Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36. doi:10.1093/jpepsy/27.1.27
- Christensen, P., & James, A. (Eds.). (2000). *Research with children: Perspectives and practices*. New York: Falmer Press.
- Finlay, L., & Gough, B. (Eds.). (2003). *Reflexivity: A practical guide for researchers in health and social sciences*. Oxford, England: Blackwell.
- Hashway, R. M. (1978). *Objective mental measurement*. New York: Praeger.
- Henry, A. (2000). *The Pediatric Interest Profiles: Surveys of play for children and adolescents*. San Antonio, TX: Therapy Skill Builders.
- James, A., Jenks, C., & Prout, A. (1998). *Theorizing childhood*. New York: Teachers College Press.
- Keller, J., Kafkes, A., Basu, S., Federico, J., & Kielhofner, G. (2005). *The Child Occupational Self Assessment (COSA)* (Version 2.1). Chicago: University of Illinois, College of Applied Health Sciences, Department of Occupational Therapy, MOHO Clearinghouse.
- Keller, J., Kafkes, A., & Kielhofner, G. (2005). Psychometric characteristics of the Child Occupational Self Assessment (COSA), part one: An initial examination of psychometric properties. *Scandinavian Journal of Occupational Therapy, 12*, 118–127. doi:10.1080/11038120510031752
- Keller, J., & Kielhofner, G. (2005). Psychometric characteristics of the Child Occupational Self-Assessment (COSA), part two: Refining the psychometric properties. *Scandinavian Journal of Occupational Therapy, 12*, 147–158. doi:10.1080/11038120510031761
- Kielhofner, G. (2008). *A Model of Human Occupation: Theory and application* (4th ed.). Baltimore: Lippincott Williams & Wilkins.
- Kielhofner, G., Forsyth, K., Kramer, J., & Iyenger, A. (2009). Developing the Occupational Self Assessment: The use of Rasch analysis to assure internal validity, sensitivity, and reliability. *British Journal of Occupational Therapy, 72*, 94–104.
- King, G., Law, M., King, S., Hanna, S., Kertoy, M., Rosenbaum, P., et al. (2004). *Children's Assessment of Participation and Enjoyment (CAPE) and Preferences for Activities of Children (PAC)*. San Antonio, TX: Harcourt Assessment.
- Law, M., Baptiste, S., Carswell, A., McColl, M. A., Polatajko, H., & Pollock, N. (1998). *Canadian Occupational Performance Measure* (3rd ed.). Toronto, Ontario: Canadian Association of Occupational Therapists.
- Linacre, J. M. (2004). *Winsteps* (Version 3.49). Author.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–174. doi:10.1007/BF02296272
- 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. doi:10.1037/0003-066X.50.9.741
- Missiuna, C., Pollock, N., & Law, M. (2004). *Perceived Efficacy and Goal Setting System (PEGS)*. San Antonio, TX: Psychological Corporation.
- Rapley, M. (2004). *The social construction of intellectual disability*. Cambridge, England: Cambridge University Press.
- Shields, B. J., Palermo, T. M., Powers, J. D., Fernandez, S. A., & Smith, G. A. (2005). The role of developmental and contextual factors in predicting children's use of a visual analogue scale. *Children's Health Care, 34*, 273–287. doi:10.1207/s15326888chc3404\_3
- Smith, E. V., Jr. (2001). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. *Journal of Applied Measurement, 2*, 281–311.
- Smith, E. V., Jr. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*, 205–231.
- Smith, E. V., Jr. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith, Jr., & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp. 93–122). Maple Grove, MN: JAM Press.
- Veloza, C., Kielhofner, G., Gern, A., Lin, F., Azhar, F., Lai, J. S., et al. (1999). Worker Role Interview: Toward validation of a psychosocial work-related measure. *Journal of Occupational Rehabilitation, 9*, 153–168. doi:10.1023/A:1021397600383
- Wolfe, E. W., & Smith, E. V. Jr. (2007). Instrument development tools and activities for measure validation using Rasch models: Part II—Validation activities. *Journal of Applied Measurement, 8*, 204–234.
- Wright, B. D., & Linacre, M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 70*, 857–860.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago: MESA Press.
- Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.