Teacher Education and Special Education
2018, Vol. 41(1) 58–69

DOI: 10.1177/0888406416687814
journals.sagepub.com/home/tese
SAGE

Establishing Content Validity of the Quality Indicators for Classrooms Serving Students With Autism Spectrum Disorders Instrument

Cynthia E. Pearl[1], Eleazar Vasquez III[1], Matthew T. Marino[1], Wilfred Wienke[1], Claire Donehower[1], Jillian Gourwitz[1], Michael S. Rosenberg[2], and Sunny R. Duerr[2]

Abstract

The purpose of this study was to provide content validation for quality indicators included in an observation instrument developed to evaluate classrooms serving students with autism spectrum disorders (ASD). A panel of 103 subject-matter experts consisting of a mixture of field personnel (*n* = 64; 59 classroom teachers, five school administrators) and university faculty (*n* = 39) provided feedback regarding construct validity. Results were analyzed using two approaches: a simple aggregation of responses in the form of averages, with an a priori threshold for the determination of what will be considered acceptable, and Lawshe's Content Validity Ratio methodology. All items in the instrument were deemed to have demonstrated content validity as did the overall instrument. Based on the data from this study, the Quality Indicators for Classrooms Serving Students With ASD (QIASD) instrument can be considered to have demonstrated content validity.

Keywords

autism, accountability, teacher preparation policy/service delivery, assessment, teacher preparation practices and outcomes

Education policy has long focused on accountability of schools to meet academic standards. The way accountability is measured has, however, evolved and the tools and procedures for measuring effective teaching have varied over time and across and within states (Darling-Hammond, Wise, & Pease, 1983; Doherty & Jacobs, 2015; Doyle, 1977; Goe, Bell, & Little, 2008; Harris, Ingle, & Rutledge, 2014). Recently, the focus has shifted from teacher qualifications (i.e., "highly qualified" teacher status) to teacher effectiveness (e.g., measuring student outcomes). Previously, state efforts focused on implementation of the guidelines and standards for credentialing of special education teachers imposed by No Child Left Behind Act (NCLB, 2002) and Individuals With Disabilities Education Act (IDEA). Now, state plans align with the current legislative emphasis on accountability for student outcomes (U.S. Department of Education, Office of Planning (2015), Evaluation and Policy Development, 2009).

[1]University of Central Florida, Orlando, USA
[2]State University of New York at New Paltz, USA

Corresponding Author:
Eleazar Vasquez III, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-1250, USA.
Email: eleazar.vasquez@ucf.edu

As teacher evaluation systems emerged, the field of special education responded with attention focused on the appropriateness of alternative measures including value-added approaches for determining the effectiveness of special education teachers (Buzick & Jones, 2015; Kearns, Kleinert, Thurlow, Gong, & Quenemoen, 2015; Johnson & Semmelroth, 2014a; Jones & Brownell, 2014; Woolf, 2015). The Council for Exceptional Children (CEC, 2012) released its *Position on Special Education Teacher Evaluation* which identified five major components of an effective teacher evaluation system for special educators. Specifically, the teacher evaluation system shall (a) include fundamental system-wide components, (b) identify the complex role of the special education teacher, (c) measure the use of evidence-based practices (EBPs), (d) recognize the professionalism of special education teachers, and (e) continually incorporate findings from the research. CEC (2012) highlighted the complexity of evaluating special educators by placing emphasis on the need to consider the specific roles and responsibilities of a given special educator as well the range of exceptionalities of his or her students. CEC further maintained special education teacher evaluations are only effective if they include accurate and reliable indicators of special education teacher contributions to student growth, promote teaching as a profession, and address the persistent interrelated problems of special education teacher attrition and retention.

Over the past 5 years, the relative merits of value-added models for teacher evaluation, including those for special educators, have increasingly been questioned (American Statistical Association, 2014; Gansle et al., 2015; Harris & Herrington, 2015). Consequently, the recently reauthorized Every Student Succeeds Act (ESSA; 2015) not only maintains a focus on accountability systems and teacher effectiveness but also provides increased flexibility, placing the responsibility for developing and implementing teacher evaluation systems in the hands of states and local education agencies (LEAs). ESSA does provide federal funds via the Teacher and School Leader Incentive Program. The purposes are to support state and district innovation to "develop, implement, improve, or expand comprehensive performance-based compensation systems or human capital management systems for teachers, principals, or other school leaders . . . who raise student achievement" (Section 2211 (a)(1)) and "to evaluate the effectiveness, fairness, quality, consistency, and reliability of the systems" (Section 2211 (a)(2)).

States and LEAs are provided the opportunity to restructure teacher evaluation systems and to develop more individualized special education teacher evaluations. Researchers have delineated the challenges associated with special education teacher evaluation (Buzick & Jones, 2015; CEC, 2012; Kearns et al., 2015; Johnson & Semmelroth, 2014b; Jones & Brownell, 2014; Woolf, 2015). Some states have recognized the need for teacher evaluation systems that incorporate individualized approaches. The District of Columbia Public Schools' (2014) evaluation system, IMPACT, is one example of a teacher evaluation process that includes modifications based on the specific roles and responsibilities of special educators. The 2015-2016 IMPACT Guidebooks recognize five categories of special education (e.g., special education teachers, special education teachers—autism program, special education teachers—early childhood education, special education coordinators, and related services providers). The Massachusetts Model System for Educator Evaluation also allows districts to adapt indicators based on teacher roles and includes rubrics designed for teachers who work with specific special populations (Massachusetts Department of Elementary and Secondary Education, 2012).

Clearly, one of the most critical issues facing the field of special education is the identification of methods to evaluate special education teachers in fair and valid ways, leading to increased special educator retention and ultimately better outcomes for students with disabilities (Kearns et al., 2015). According to Johnson and Semmelroth (2014b),

> Through the implementation of a teacher evaluation system that both informs and evaluates special education teachers based on their effective use of evidence-based instructional practices, practitioners will be provided ongoing opportunities to maintain relevancy, practicality, and applicability of their work. (p. 80)

A classroom observation conducted by a supervisor or school administrator is the most widely used measure of teacher effectiveness and may be "used to measure observable classroom processes, including specific teacher practices, holistic aspects of instruction, and interactions between teachers and students" (Goe et al., 2008, p. 16). Given the problems associated with application of value-added measures to special education teachers (Ballou & Springer, 2015; Darling-Hammond, 2015; Harris & Herrington, 2015; Gansle et al., 2015; Goldhaber, 2015; Jones, Buzick, & Turkan, 2013), it can be anticipated that increased emphasis will be placed on observation systems as a data source for teacher evaluation (Jones & Brownell, 2014). A range of teacher observation systems are available and effective in discriminating between effective teachers and those who are less effective (Klinger, Brownell, & Bateman, 2011). Jones and Brownell (2014) examined the potential for validating Charlotte Danielson's commonly used Framework for Teaching (FFT) as an observation system for special education teachers and concluded there is a need for considerable research. Identified challenges included lack of awareness of effective practices for teaching students with disabilities on the part of administrators conducting observations and lack of assessment of collaboration with colleagues deemed an integral part of special educator efficacy.

Over the past decade, researchers have gone to considerable efforts to set criteria for identifying specific teaching practices as evidence based (Horner et al., 2005; Nathan & Gorman, 2002; Odom et al., 2005). Subsequent application of these criteria has resulted in the identification of a number of EBPs specific to working with students with disabilities. Given the exponential increase in the number of students identified with autism spectrum disorders (ASD; Centers for Disease Control and Prevention, 2014), particular attention has been paid to identifying EBPs specific to the ASD population. A number of researchers and consortiums have accumulated lists of EBPs for working with students with ASD (National Autism Center [NAC], 2009; Rogers & Vismara, 2008; Simpson, 2005; Wong et al., 2015).

Unfortunately, despite the strong evidence-base for a number of practices for teaching students with ASD, researchers have found many teachers lack preparation and support for the implementation of those EBPs (Belfiore, Fritts, & Herman, 2008; Brock, Huber, Carter, Juarez, & Warren, 2014; Browder & Cooper-Duffy, 2003; National Research Council [NRC], 2001). Project ASD's Quality Indicators for Classrooms Serving Students With ASD (QIASD) is an observational tool designed specifically to provide special education teachers serving students with ASD with discrete and actionable feedback (Johnson & Semmelroth, 2014a). In this study, we provide content validation for the 52 quality indicators assessed by the QIASD. Goe and colleagues (2008) advised that when employing classroom observation as a teacher evaluation method, careful attention must be paid to the validity and reliability of protocols. They noted that when assessing the validity of classroom observation instruments, it is essential to account for "the instrument's ability to measure how well a teacher exemplifies standards of practice that have been deemed important for that grade level, subject, and teaching context by some group of experts" (Goe et al., 2008, p. 50).

Method

One common practice for establishing content validity is to employ the judgment of subject-matter experts and use the resulting feedback to provide evidence supporting claims of validity. If the panel endorses an item, it can be considered to have construct validity, whereas the panel's rejection of an item can lead to the item being either discarded or rewritten. Although there are several methodological approaches to analyzing the resulting data in such investigations, we provide two: (a) a simple aggregation of responses in the form of averages, with an a priori threshold for the determination of what will be considered acceptable, and (b) a method established by Lawshe (1975) that is more rigorous and grounded in statistics. Both approaches evaluate the degree to which any individual item is endorsed by the panel of experts and provide aggregated indexes.

Lawshe (1975) viewed content validity as the degree to which a content evaluation panel perceives overlap between an instrument's items and the domain those items are designed to represent. Specifically, he argued that people who are embedded within the domain, or "those who 'know the job' are normally competent to make the required judgments" (Lawshe, 1975, p. 566). In addition, Lawshe recommended that the panel be comprised of both "incumbents and supervisors" (p. 566), which in the context of the present study is operationalized as the inclusion of experts, both higher education faculty and practicing special educators in the areas of ASD and special education.

Participants

The panel of experts included 103 participants consisting of a combination of practicing field personnel (*n* = 64; 59 special education classroom teachers and five school administrators) and university faculty (*n* = 39). University faculty were identified through email contact of special education unit heads at research universities, requesting recommendations of faculty with expertise in ASD and experience in preparing teachers to serve individuals with ASD. Forty-seven heads from research universities recommended faculty members from their institutions. Those faculty recommended for inclusion were contacted by email to verify their expertise in ASD and to request their participation in the validation study. Faculty contacted were also invited to assist with the identification of expert field personnel. Specifically, they were asked to share the survey link with fully certified special education teachers and administrators, currently serving students with ASD, who they deemed to be highly effective educators. In return for completing the survey, all participants were given the option of having their names entered into a drawing with the chance to win one of five Amazon gift cards, valued at US$50 each.

Instrument and Procedure

Participants were asked to respond to several items within a Qualtrics survey. They were asked to rate each item on the QIASD evaluation tool, indicating the degree to which the panelist agreed or disagreed with the item's alignment with the seven specific CEC standards. Responses were on a 5-point Likert-type scale, ranging from *strongly disagree* to *strongly agree*. In addition to these scale responses, participants were given the opportunity to provide open-ended feedback at the end of each CEC standard to present any additional indicators that might enhance the evaluation of the construct.

The QIASD was developed as a product of Project ASD, funded through the Office of Special Education Programs (OSEP) and located at the University of Central Florida. It was designed to guide a classroom observer in evaluating the strength and consistency of specific indicators of quality educational programming for students with ASD. It includes quality indicators from the Observation Assessment for Classrooms Serving Students With Autism Spectrum Disorders (OAASD), developed as a product of a PEPSA (Partnership for Effective Programs for Students with Autism) and subsequently revised and adopted by Florida Center for Autism Related Disorders (CARD) Centers.

The QIASD reflects revisions to quality indicators based on field testing of the OAASD as well as additions based on an extensive review of the literature. The QIASD also aligns with the seven Initial Preparation Standards and the Initial Specialty Set: Developmental Disabilities and Autism Spectrum Disorder developed by the CEC. The specialty set standards capture the professional knowledge base, including empirical research, disciplined inquiry, informed theory, and the wisdom of practice for their area of expertise for each proposed knowledge and skill (CEC, 2015).

The QIASD consists of 52 quality indicators aligned with the seven CEC standards: (a) learner development and individual learning differences, (b) learning environments, (c) instruction curricular content knowledge, (d) assessment, (e) instructional planning and strategies, (f) professional learning and practice, and (g) collaboration. Each indicator is given a score of 0 to 4 or NA. Quality indicators receive a 0 if unsatisfactory (not present), 1 if developing (very limited presence), 2 if needs improvement (somewhat present), 3 if effective (present), 4 if highly effective (very much present), and NA (unrated) if there was not an opportunity to observe the quality indicator during the 1-hour observation. In addition, a 13-item interview protocol addressed specific indicators observers may not have an opportunity to observe while in the classroom (e.g., family training sessions, family involvement in Individualized Education Program [IEP] meetings).

Analysis

Descriptive statistics. The mean and standard deviation for each item were calculated for field personnel and for university faculty, as well as the overall mean and standard deviation for all participants. Any individual item with a mean response rate below 4.25 was flagged for additional consideration. In addition, mean responses for each item were compared for group differences between field personnel and university faculty using Student's *t* test, providing additional depth in the analysis of response patterns. Aggregated mean ratings were calculated by standard and for the overall instrument.

Content Validity Ratio (CVR). In its original form, Lawshe's (1975) approach to establishing content validity asks participants to rate each item on a 3-point scale, where each item is judged to be either *essential*, *important but not essential*, or *unimportant*. The number of *essential* responses is then counted, and a CVR of the number of essential responses to the total number of responses is calculated using the following formula:

$$\text{CVR} = \frac{n_e - \frac{N}{2}}{\frac{N}{2}}.$$

In this formula, n_e represents the number of panelists to rate the item as being *essential*, and N is the total number of responses for the item. The resulting CVR is effectively a ratio of *essential* endorsements to the total number of responses. The value for CVR will range between −1 and +1, and if the CVR is positive, it is because more participants rated the item as being *essential* than not. The greater the CVR (i.e., the closer CVR is to 1), the more agreement there was among the panelists, with a value of 1 representing perfect agreement. A negative CVR demonstrates that the majority of panelists do not see the item as being *essential*, and the researcher should discard or rewrite the item. In addition, guidelines have been provided for acceptable CVR values based on the number of participants (Ayre & Scally, 2014; Lawshe, 1975), effectively establishing thresholds beyond which a CVR can be considered to demonstrate levels of agreement above those of chance. Based on these minimum values of CVR and a sample size of 30, an acceptable CVR rating would be 0.33, which we use as an acceptable approximation for the present study because the smaller pool of participants, university faculty, had 39 responses. A larger number of participants makes the minimum value for the CVR decrease, so using 0.33 can be considered a more conservative threshold.

Once the CVR has been calculated for each item and items have been either revised or discarded based upon the CVR, an index of the remaining items can be calculated by averaging the CVR for all remaining items. It is important to note that revised or new items would also need to be validated, and a CVR would need to be calculated for these items as well. The resulting index, referred to as the

Content Validity Index (CVI), provides a quantitative value for the overall measure, with CVI values being interpreted similar to the way individual CVR values are interpreted (i.e., a measure with a positive CVI value can be considered to have demonstrated content validity, with values closer to 1 indicating stronger evidence of content validity).

For the present validation effort, we adapted Lawshe's approach (see Johnston & Wilkinson, 2009). Instead of being asked to rate each item on a 3-point scale, participants were asked to rate each item on how well it aligned with the CEC standards using a traditional 5-point Likert-type scale ranging from *strongly disagree* to *strongly agree*. To adapt the resulting data to the Lawshe paradigm, we consider all responses of either *agree* or *strongly agree* (i.e., a rating of 4 or 5) as being an endorsement of the item's alignment with the corresponding CEC standard. Thus, the resulting CVR values represent the ratio of participants who endorsed the item with at least a response of *agree* over the total number of responses for the item. In this context, any item or index with a CVR/CVI value that is deemed to be acceptably large (i.e., any value that exceeds 0.33) can be considered to have been endorsed by the collective panel of experts, with values closer to 1 representing a stronger overall endorsement.

Missing data. Missing data is a common issue in data analysis, and the ways in which missing data are handled are plentiful and varied; common methods include listwise deletion (removing the entire participant from the data analysis) and mean substitution (replacing the missing data with the mean response for the item). Within the present context, however, items are being analyzed individually, and so missing data is less problematic. The CVR for items in which a participant was missing a response was calculated using the same formula, because the missing value would affect both the numerator of the CVR equation and the denominator, effectively making the missing value moot. That said, there was very little missing data in this response set, with 94.4% of the fields containing data.

Results

Descriptive Statistics

Means and standard deviations were calculated for each item, both by group membership (field personnel and university faculty) and overall. These values are presented in Table 1. For each item, values with a mean response of less than 4.25 were identified for additional consideration, and items with a mean of 4.25 or higher were considered to be highly endorsed by the panel of experts. For the field personnel, Items 2.c, 5.b, 5.c, and 5.e required additional consideration, and for the university faculty, Items 2.c and 5.b required additional consideration. When the overall mean was considered, three of these items (2.c, 5.b, and 5.c) had average ratings below this threshold (3.98, 4.14, and 4.21, respectively).

Aggregated mean ratings for the items in each standard were calculated and are reported in Table 2. These data demonstrate on the original scale of 1 to 5, participants' endorsement of the alignment between the items and the CEC standards is strong, with mean ratings greater than 4.5 on the entire instrument for both field personnel and university faculty and the lowest mean rating for any given standard's items being greater than 4.5 as well; Standard 2's mean ratings were between 4.2 and 5.4.

Group Differences in Responses

To provide additional depth to the analysis of the response pattern, Student's t value was calculated to determine whether the average difference in ratings between field personnel and university faculty for each item was large enough to be considered statistically significant. For the purpose of this analysis, the calculated value for Student's t, reported in the final column of Table 1, would need to exceed an absolute value of 1.98 ($\alpha = .05$). This proved to be the case for 53% of the items. In all cases where the difference was statistically significant, the average rating from university faculty was higher than the average rating provided by field personnel, indicating that university faculty perceived stronger alignment with the CEC standards on most items.

Table 1. Average Ratings by Item by Role.

Item number	Field personnel *M*	Field personnel *SD*	Field personnel *n*	University faculty *M*	University faculty *SD*	University faculty *n*	Aggregated *M*	Aggregated *SD*	*t* value
1.a	4.53	0.85	64	4.90	0.31	39	4.67	0.72	**−2.574**
1.b	4.48	0.85	64	4.77	0.43	39	4.59	0.73	−1.937
1.c	4.47	0.82	64	4.72	0.51	39	4.56	0.72	−1.712
1.d	4.36	1.01	64	4.62	0.59	39	4.46	0.88	−1.434
1.e	4.36	0.90	64	4.77	0.54	39	4.51	0.80	**−2.582**
1.f	4.58	0.81	64	4.84	0.44	38	4.68	0.71	−1.848
2.a	4.47	0.80	62	4.39	0.82	38	4.44	0.81	0.437
2.b	4.60	0.80	62	4.74	0.50	38	4.65	0.70	−0.969
2.c	**4.08**	1.26	62	**3.82**	1.11	38	**3.98**	1.21	1.067
2.d	4.39	0.95	62	4.53	0.86	38	4.44	0.91	−0.738
2.e	4.32	1.04	62	4.61	0.68	38	4.43	0.92	−1.494
2.f	4.56	0.84	62	4.71	0.57	38	4.62	0.75	−0.946
2.g	4.58	0.80	62	4.61	0.64	38	4.59	0.74	−0.161
2.h	4.42	0.95	62	4.74	0.50	38	4.54	0.82	−1.900
2.i	4.51	0.87	61	4.63	0.49	38	4.56	0.75	−0.800
2.j	4.29	1.05	62	4.32	0.84	38	4.30	0.97	−0.127
2.k	4.56	0.72	61	4.78	0.53	37	4.64	0.66	−1.656
3.a	4.40	0.79	58	4.76	0.43	38	4.54	0.69	**−2.605**
3.b	4.36	0.85	58	4.53	0.80	38	4.43	0.83	−0.947
3.c	4.50	0.71	58	4.87	0.34	38	4.65	0.62	**−2.987**
3.d	4.47	0.73	58	4.84	0.37	38	4.61	0.64	**−2.937**
3.e	4.38	0.81	58	4.71	0.52	38	4.51	0.73	**−2.233**
3.f	4.51	0.76	57	4.92	0.27	38	4.67	0.64	**−3.209**
3.g	4.32	0.87	57	4.76	0.49	38	4.49	0.77	**−2.879**
4.a	4.53	0.71	58	4.87	0.41	38	4.67	0.63	**−2.631**
4.b	4.47	0.75	58	4.89	0.31	38	4.64	0.65	**−3.323**
4.c	4.55	0.68	58	4.87	0.34	38	4.68	0.59	**−2.656**
4.d	4.57	0.70	58	4.79	0.47	38	4.66	0.63	−1.695
5.a	4.50	0.71	58	4.97	0.16	38	4.69	0.60	**−4.053**
5.b	**4.16**	0.89	58	**4.11**	1.13	38	**4.14**	0.99	0.240
5.c	**4.17**	1.03	58	4.26	1.08	38	**4.21**	1.05	−0.414
5.d	4.34	0.81	58	4.66	0.81	38	4.47	0.82	−1.852
5.e	**4.24**	0.96	58	4.53	1.06	38	4.35	1.01	−1.365
5.f	4.60	0.67	58	4.79	0.41	38	4.68	0.59	−1.523
5.g	4.35	0.88	57	4.68	0.53	37	4.48	0.77	**−2.025**
5.h	4.55	0.71	58	4.53	0.83	38	4.54	0.75	0.161
5.i	4.66	0.69	58	4.82	0.39	38	4.72	0.59	−1.303
5.j	4.62	0.67	58	4.95	0.23	38	4.75	0.56	**−2.891**
5.k	4.53	0.71	58	4.82	0.46	38	4.65	0.63	**−2.174**
5.l	4.37	0.98	57	4.82	0.46	38	4.55	0.83	**−2.638**
5.m	4.38	0.91	55	4.82	0.39	38	4.56	0.77	**−2.756**
6.a	4.47	0.84	58	4.30	0.85	37	4.40	0.84	0.948
6.b	4.59	0.70	58	4.89	0.31	37	4.71	0.60	**−2.491**
6.c	4.59	0.80	58	4.59	0.96	37	4.59	0.86	−0.046
7.a	4.29	0.90	58	4.54	0.61	37	4.39	0.80	−1.474
7.b	4.48	0.75	58	4.51	0.77	37	4.49	0.76	−0.192

(continued)

Table 1. (continued)

Item number	Field personnel			University faculty			Aggregated		
	M	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*	*t* value
7.c	4.59	0.73	58	4.86	0.35	37	4.69	0.62	**−2.178**
7.d	4.57	0.75	58	4.86	0.35	37	4.68	0.64	**−2.243**
7.e	4.47	0.78	58	4.81	0.46	37	4.60	0.69	**−2.439**
7.f	4.34	0.91	58	4.92	0.28	37	4.57	0.78	**−3.726**
7.g	4.36	0.91	58	4.89	0.31	37	4.57	0.78	**−3.402**

Note. t value is the value for Student's *t* for the mean difference between field personnel and university faculty; bold *t* values indicate statistically significant differences.

Table 2. Aggregated Average Rating per Standard by Role.

	Overall		Field personnel		University faculty	
	M	*SD*	*M*	*SD*	*M*	*SD*
Overall	4.54	0.76	4.45	0.83	4.69	0.55
CEC Standard 1	4.58	0.76	4.46	0.87	4.77	0.47
CEC Standard 2	4.47	0.84	4.43	0.92	4.53	0.69
CEC Standard 3	4.56	0.70	4.42	0.79	4.77	0.46
CEC Standard 4	4.66	0.62	4.53	0.71	4.86	0.39
CEC Standard 5	4.52	0.77	4.42	0.82	4.67	0.61
CEC Standard 6	4.56	0.77	4.55	0.78	4.59	0.71
CEC Standard 7	4.57	0.72	4.44	0.82	4.77	0.45

Note. CEC = Council for Exceptional Children.

CVR

The CVR was calculated for each item using the methodology previously described, and the resulting CVR values for each item are reported by participant role and in aggregated form for the overall sample (see Table 3). In the aggregated form, all items exceeded the CVR threshold of 0.33, with the lowest CVR value belonging to Item 2.c (CVR = 0.48). For the disaggregated data, Item 2.c was the lowest rated item for both field personnel (CVR = 0.61) and university faculty (CVR = 0.26). In fact, the latter is the only instance where the calculated CVR was lower than the threshold of 0.33, but the corresponding CVR for field personnel on this item affected the overall CVR rating. Based on the previously discussed threshold for CVR, all the items on the instrument have demonstrated content validity, with Item 2.c being the only item that may need to be considered for revision or removal.

CVI

Having determined all items have strong content validity, a CVI was calculated for each section of the instrument, corresponding with each CEC standard; an overall CVI for the instrument was also calculated. As Table 4 shows, each subsection can be said to possess content validity and the CVI values indicate a level of agreement among raters that could be described as very strong. The lowest CVI was for the section corresponding to CEC Standard 2, which had a CVI of 0.804, while the overall instrument CVI was 0.857. These values are significantly higher than the necessary CVI threshold of 0.33.

Table 3. Content Validity Ratios by Item by Role.

Item number	Field personnel	University faculty	Combined
1.a	0.91	1.00	0.94
1.b	0.91	1.00	0.94
1.c	0.91	0.95	0.92
1.d	0.75	0.90	0.81
1.e	0.81	0.90	0.84
1.f	0.91	0.95	0.90
2.a	0.94	0.68	0.80
2.b	0.94	0.95	0.94
2.c	0.61	0.26	0.48
2.d	0.81	0.79	0.80
2.e	0.74	0.79	0.76
2.f	0.87	0.89	0.88
2.g	0.94	0.84	0.90
2.h	0.81	0.95	0.86
2.i	0.90	1.00	0.92
2.j	0.68	0.53	0.64
2.k	0.90	0.89	0.86
3.a	0.86	1.00	0.88
3.b	0.76	0.74	0.75
3.c	0.93	1.00	0.96
3.d	0.90	1.00	0.94
3.e	0.76	0.95	0.83
3.f	0.93	1.00	0.94
3.g	0.72	0.95	0.81
4.a	0.93	0.95	0.96
4.b	0.86	1.00	0.92
4.c	0.97	1.00	0.98
4.d	0.93	0.95	0.94
5.a	0.93	1.00	0.96
5.b	0.72	0.47	0.63
5.c	0.76	0.47	0.65
5.d	0.83	0.79	0.81
5.e	0.69	0.68	0.69
5.f	0.97	1.00	0.98
5.g	0.82	0.95	0.83
5.h	0.93	0.79	0.91
5.i	0.93	1.00	0.96
5.j	0.97	1.00	0.98
5.k	0.93	0.95	0.94
5.l	0.68	0.95	0.77
5.m	0.78	1.00	0.83
6.a	0.79	0.62	0.76
6.b	0.93	1.00	0.96
6.c	0.86	0.89	0.87
7.a	0.72	0.89	0.79
7.b	0.86	0.78	0.83
7.c	0.90	1.00	0.94
7.d	0.86	1.00	0.92
7.e	0.83	0.95	0.87
7.f	0.76	1.00	0.85
7.g	0.83	1.00	0.89

Note. t value is the value for Student's *t* for the mean difference between field personnel and university faculty; bold *t* values indicate statistically significant differences.

Table 4. Content Validity Indexes for Each Subsection by Role.

	Field personnel	University faculty	Combined
Overall instrument	0.846	0.882	0.857
CEC Standard 1	0.865	0.948	0.893
CEC Standard 2	0.830	0.780	0.804
CEC Standard 3	0.837	0.947	0.872
CEC Standard 4	0.922	0.974	0.948
CEC Standard 5	0.842	0.850	0.841
CEC Standard 6	0.862	0.838	0.865
CEC Standard 7	0.823	0.946	0.871

Note. CEC = Council for Exceptional Children.

Conclusion and Future Directions

Project ASD's QIASD is an observational tool specifically designed to support special education teachers serving students with ASD. The objective of this study was to establish the content validity of the instrument. A panel of 103 experts, representing university professors and field personnel, was asked to assess the alignment between each item and its corresponding CEC standard. Descriptive statistics were calculated, and data were analyzed using Lawshe's (1975) CVR methodology. All items in the instrument were deemed to have demonstrated content validity, and each subsection of the instrument had a very high CVI, as did the overall instrument.

Based on the data from this study, the QIASD instrument can be considered to have content validity. The reauthorized ESSA (2015) continues a focus on accountability systems and teacher effectiveness providing increased flexibility by placing responsibility for teacher evaluation systems in the hands of states and LEAs. The results from this study add to the body of literature by providing a preliminary first step in the development of a validated instrument for assessing teacher performance with students with ASD. Given the functional form of this instrument, additional analysis and research should be conducted to include psychometric analyses centering on construct validity and reliability. In addition, future research associated with Project ASD and similar efforts could develop interventions impacting teacher performance utilizing this instrument to assess efficacy.

Authors' Note

The contents of this article do not necessarily represent the policy of the U.S. Department of Education, and readers should not assume endorsement by the Federal Government (Project Officer, Celia Rosenquist).

Declaration of Conflicting Interests

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

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The contents of this article were developed under a grant from the U.S. Department of Education H325K130222.

References

American Statistical Association. (2014, April). *ASA statement on using value-added models for educational assessment*. Retrieved from https://www.amstat.org/asa/files/pdfs/POL-ASAVAM-Statement.pdf

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's Content Validity Ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, *47*, 79-86.

Ballou, D., & Springer, M. G. (2015). Using student test scores to measure teacher performance: Some problems in the design and implementation of evaluation systems [Special issue]. *Educational Researcher*, *44*, 77-86. doi :10.3102/0013189X15574904

Belfiore, P. J., Fritts, K. M., & Herman, B. C. (2008). The role of procedural integrity using self-monitoring to enhance discrete trial instruction (DTI). *Focus on Autism and Other Developmental Disabilities*, *23*, 95-102.

Brock, M. E., Huber, H. B., Carter, E. W., Juarez, A. P., & Warren, Z. E. (2014). Statewide assessment of professional development needs related to educating students with autism spectrum disorder. *Focus on Autism and Other Developmental Disabilities*, *29*, 67-79.

Browder, D., & Cooper-Duffy, K. (2003). Evidence-based practices for students with severe disabilities and the requirements for accountability in No Child Left Behind. *The Journal of Special Education*, *37*, 157-163.

Buzick, H. M., & Jones, N. D. (2015). Using test scores from students with disabilities in teacher evaluation. *Educational Measurement*, *34*(3), 28-38.

Centers for Disease Control and Prevention. (2014). *Prevalence of autism spectrum disorder among children ages 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2010*. Retrieved from http://www.cdc.gov/mmwr/preview/mmwrhtml/ss6302a1.htm?s_cid=ss6302a1_w

Council for Exceptional Children. (2012). *CEC position on special education teacher evaluation*. Arlington, VA: Author.

Council for Exceptional Children. (2015). *What every special educator must know: Professional ethics and standards* (7th ed.). Retrieved from http://www.cec.sped.org/standards/

Darling-Hammond, L. (2015). Can value added add value to teacher evaluation? *Educational Researcher*, *44*, 132-137. doi:10.3102/0013189X15575346

Darling-Hammond, L., Wise, A. E., & Pease, S. R. (1983). Teacher evaluation in the organizational context: A review of the literature. *Review of Educational Research*, *53*, 285-328.

District of Columbia Public Schools. (2014). *Impact: The District of Columbia Public Schools Effectiveness Assessment System for School-Based Personnel*. Retrieved from http://dcps.dc.gov/node/979742

Doherty, K. M., & Jacobs, S. (2015, November). *State of the states: Evaluating teaching, leading and learning, National Council on Teacher Quality*. Retrieved from http://www.nctq.org/dmsView/StateofStates2015#page=3&;zoom=auto,-194,80

Doyle, W. (1977). Paradigms for research on teacher effectiveness. *Review of Research in Education*, *5*, 163-198.

Every Student Succeeds Act. (2015). Retrieved from https://www.congress.gov/bill/114th-congress/senate-bill/1177/text#toc-HB1FE71BC-6C094FA285180CCC608ADADE

Gansle, K. A., Noell, G. H., Grandstaff-Beckers, G., Stringer, A., Roberts, N., & Burns, J. M. (2015). Value-added assessment of teacher preparation: Implications for special education. *Intervention in School and Clinic*, *51*, 106-111.

Goe, L., Bell, C., & Little, O. (2008). Approaches to Evaluating Teacher Effectiveness: A Research Synthesis. *National Comprehensive Center for Teacher Quality*.

Goldhaber, D. (2015). Exploring the potential of value-added performance measures to affect the quality of the teacher workforce. *Educational Researcher*, *44*, 87-95.

Harris, D. N., & Herrington, C. D. (2015). Editors' introduction: The use of teacher value-added measures in schools: New evidence, unanswered questions, and future prospects. *Educational Researcher*, *44*, 71-76.

Harris, D. N., Ingle, W. K., & Rutledge, S. A. (2014). How teacher evaluation methods matter for accountability: A comparative analysis of teacher effectiveness ratings by principles and teacher value-added measures. *American Educational Research Journal*, *51*, 73-112.

Horner, R., Carr, E., Halle, J., Mcgee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*, 165-179.

Johnson, E., & Semmelroth, C. L. (2014a). Introduction to AEI's special issue on special education teacher evaluations. *Assessment for Effective Intervention*, *39*, 67-70.

Johnson, E., & Semmelroth, C. L. (2014b). Special education teacher evaluation: Why it matters, what makes it challenging, and how to address these challenges. *Assessment for Effective Intervention*, *39*, 71-82.

Johnston, P., & Wilkinson, K. (2009). Enhancing validity of critical tasks selected for college and university program portfolios. *National Forum of Teacher Education Journal*, *19*(3), 1-6.

Jones, N. D., & Brownell, M. T. (2014). Examining the use of classroom observations in the evaluation of special education teachers. *Assessment for Effective Intervention*, *39*, 112-124.

Jones, N. D., Buzick, H. M., & Turkan, S. (2013). Including students with disabilities and English learners in measures of educator effectiveness. *Educational Researcher*. doi:0013189X12468211.

Kearns, J. F., Kleinert, H. L., Thurlow, M. L., Gong, B., & Quenemoen, R. (2015). Alternate assessments as one measure of teacher effectiveness: Implications for our field. *Research and Practice for Persons With Severe Disabilities*, *40*, 20-35.

Klinger, J., Brownell, M., & Bateman, D. (2011). Special education teacher evaluation. *New Times for Division for Learning Disabilities*, *29*(3), 1-5.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personality Psychology*, *28*, 563-575.

Massachusetts Department of Elementary and Secondary Education. (2012). *Massachusetts model system for educator evaluation: Part III: Guide to rubrics and model rubrics for superintendent, administrator and teacher*. Retrieved from http://www.doe.mass.edu/edeval/model/

Nathan, P., & Gorman, J. (2002). *A guide to treatments that work* (2nd ed.). New York, NY: Oxford University Press.

National Autism Center (NAC). (2009). *National Standards Re- port*. Randolph, Massachusetts: National Autism Center.

National Research Council. (2001). *Educating children with autism*. Washington, DC: National Academies Press.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425. (2002).

Odom, S., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, *71*, 137-148.

Rogers, S., & Vismara, L. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child & Adolescent Psychology*, *37*, 8-38. doi:10.1080/15374410701817808

Simpson, R. (2005). Evidence-based practices and students with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *20*, 140-149.

U.S. Department of Education. (2009, November). *Race to the Top executive summary*. Retrieved from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education, Office of Elementary and Secondary Education, & Office of State Support. (2015). *Fundamental change: Innovation in America's schools under Race to the Top*. Washington, DC: U.S. Department of Education.

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2010). *A blueprint for reform: The reauthorization of the Elementary and Secondary Education Act*. Retrieved from http://www2.ed.gov/policy/elsec/leg/blueprint/

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, *45*, 1951-1966.

Woolf, S. B. (2015). Special education professional standards: How important are they in the context of teacher performance evaluation? *Teacher Education and Special Education*, *38*, 276-290.

Author Biography

Cynthia E. Pearl is Faculty Administrator at University of Central Florida.

Eleazar Vasquez III is Associate Professor at University of Central Florida.

Matthew T. Marino is Associate Professor at University of Central Florida.

Wilfred Wienke is Professor Emeritus at University of Central Florida.

Claire Donehower is Doctoral Candidate at University of Central Florida.

Jillian Gourwitz is a Project Coordinator at University of Central Florida.

Michael S. Rosenberg is a Professor and Dean of the School of Education at the State University of New York at New Paltz.

Sunny R. Duerr is Assistant Dean of Assessment and Accreditation at SUNY New Paltz.