Appendix 2.3B.10

2018 Consequential Validity Survey Results for the Oregon Extended Assessments

Abstract

Behavioral Research and Teaching (BRT) at the University of Oregon conducted the current consequential validity survey for the Oregon Department of Education (ODE). The purpose was to determine Oregon educator perceptions of the impact that implementation of the Oregon Extended Assessment (ORExt) program has upon the field in the areas of instruction and student opportunity for students with significant cognitive disabilities (SWSCD). BRT collected the information as a basis for continuous improvement efforts related to the ORExt. This is the fifth year collecting such information, with participants asked to respond to a range of quantitative and qualitative survey prompts. All participants were Qualified Assessors (QAs) and Qualified Trainers (QTs) in Oregon, and were assured of strict confidentiality, with aggregated survey results reported here to protect confidentiality.

Theoretical Background

Messick (1989) introduced consequential validity as a concept in relation to test development and usage over two decades ago. Shepard (1997) broadened the definition, incorporating positive, negative and neutral consequences, as well as intended and unintended consequences. Broadly, consequential validity is considered as separate though interconnected to other aspects of test validity (i.e., construct, content). Whether the consequences of test use are the responsibility of the test author or user (Popham, 1997), they must be carefully considered during test development and subsequent use—as test validation depends on the decision-making procedures employed in both contexts (Kane, 2001). Further, key organizations and federal peer review requires documentation of consequential validity in reporting (AERA, NCME, & APA, 1999; OESE, 2007).

The administered survey questions were framed based upon current consequential validity approaches for alternate assessments in the literature (e.g., Lane, Parke, & Stone, 1998; Wilson, 2005), and also targeted issues that were of specific value in Oregon. A thorough description of the theoretical framework behind the approach to consequential validity in this study can be found in the 2013-14 technical report (Oregon Department of Education, 2014). 2017-18 consequential validity survey results are reported below.

Methods

Participants

Responses to survey items were received from 123 participants, though some entries included no data, and thus, *n*-sizes range across survey demographic responses and items. All demographics questions were gathered toward the end of the survey, as some attrition was expected and we prioritized items to proceed in order of utility. The 123

respondents, all of whom were Qualified Assessors (QAs; *n* = 96, 78%), Qualified Trainers (QTs; *n* = 46, 22%), in the or.k12test.com database, represented about 15% of the solicited respondents in the state (10% of available QAs and 33% of available QTs). The sample was 83% female, 14% male, and 2% other. Participants reflected age groups from 20-50 and above, with most between 41-45 (17.8%) and 51 and above (39%) years of age. The sample represented all regions of the state, with most respondents residing along the North and Central I-5 Corridors—Region 2 (Portland, Beaverton, and Hillsboro areas; 42%) and Region 5 (Eugene, Corvallis, Salem areas; 39%), respectively. Participants' educational experience ranged from 0 to over 31 years, with the most respondents having 16-20 years of experience (21%). Tables 1-6 display descriptive statistics for the sample demographics, including sex, age, educational experience, region, educational role, and level of education.

Procedure

The Oregon Extended Assessment Consequential Validity Survey was distributed via the Google Forms system (https://goo.gl/forms/P9penNdG72o0tc6f1) from May 20, 2018 through June 15, 2018. The survey link was distributed via ORExt electronic mail listservs to 956 Qualified Assessors and 139 Qualified Trainers for the ORExt. Survey responses were downloaded in an Excel comma separated values file and analyzed descriptively in Excel.

In addition to the six demographic questions, the 2017-2018 Oregon Extended Assessment Consequential Validity Survey included a total of eight quantitative and three qualitative items designed to gauge the impact of the ORExt test use. The seven positively-worded quantitative items employed a four-point scale ranging from 1-4 (1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree). One quantitative item asked for a

response by inputting numerical values for hours and minutes. There were three qualitative, open-ended response questions.

Results

Quantitative and qualitative results from the 2017-2018 Oregon Extended Assessment Consequential Validity Survey are presented, respectively.

Quantitative

Because all quantitative items were positively worded questions with answer choices on an ordinal scale, the modes of the responses are reported for interpretation.

In general, ORExt QAs and QTs deemed that the ORExt test items were easy to administer and score [Item 1; *mode* = Strongly Agree or Agree (97%)]. Respondents also thought that items were accessible for students who participated in the ORExt [Item 2; *mode* = Strongly Agree or Agree (78%)]. QAs and QTs, overall, thought that the ORExt reflected the academic content their SWSCD should be learning [Item 3; *mode* = Strongly Agree or Agree (68%)]. Further, respondents believed that the performances required by ORExt items are appropriate behaviors to review to determine what their SWSCD know and can do [Item 4; *mode* = Strongly Agree or Agree (85%)]. QTs and QAs indicated a mildly positive response to having the curricula needed to teach academic skills that are aligned to the Essentialized Assessment Frameworks to their students taking ORExt [Item 5; mode = Strongly Agree or Agree (55%)]. Furthermore, respondents generally agreed that they felt confident in interpreting the scores and the relative achievement level descriptors for ORExt [Item 6; *mode* Strongly Agree or Agree (70%)].

As time spent administering assessments is a specific concern addressed in the new Every Student Succeeds Act (2015), we asked test administrators to estimate how long it

took to administer one content area assessment in ELA or Math. Of the 123 respondents, 56% recorded that a content area assessment required between one hour and two hours to administer. Nine percent of respondents recorded that the assessment took up to three hours to administer. The majority of QAs report that the assessment takes no more than one hour to administer per content area, with a few cases taking longer than an hour for test administration.

Overall, respondents had varying responses to the potential positive *educational* impacts that OREXt implementation is having on SWSCD (Items 7a to7h). In general, respondents had a mildly positive view of educational impacts of the OREXt. Respondents believed that the OREXt increased educator understanding of the academic content for SWSCD [Item 7a; *mode* = Strongly Agree or Agree (68%)] and provided new models for assessing academics for SWSCD (Item 7b; *mode* = Strongly Agree or Agree (73%)]. While respondents were neutral that OREXt provided positive impacts in the curricular and instructional approaches used with their students [Item 7c; *mode* = Strongly Agree or Agree (50%)], Many believed that it improved the manner in which classroom assessments are designed and implemented [Item 7d; *mode* = Strongly Agree or Agree (53%)]. QAs and QTs mildly disagreed that OREXt improved the learning outcomes of SWSCD [Item 7e; *mode* = Strongly Agree or Agree (48%)], nor did it increase access to the general education curriculum for those students [Item 7f; *mode* = Strongly Agree or Agree (43%)]. However, respondents were more agreeable regarding the positive impact of OREXt in developing academic goals and objectives in IEPs for SWSCD [Item 7g; *mode* = Strongly Agree or Agree (57%)], as well as improving the alignment between IEP and state standards and benchmarks [Item 7h; *mode* = Strongly Agree or Agree (57%)].

Qualitative

Participating QAs and QTs were asked to answer three open-ended survey items. The first qualitative response item asked respondents to describe what they appreciated most about the 2017-18 ORExt. Of the 123 respondents for this item, two overall aspects of the ORExt were most appreciated (in order of greatest frequency to least):

1. Efficiency of administration, such as more streamlined administration, ease of administration while maintaining tasks that reliably measure target skills, easier to give and score.
2. Overall item and test design, including one item per page, visual supports, variety of question prompts, student materials design, and accessibility for students.

The second qualitative response item asked respondents to recommend one improvement that could be made to the 2017-18 ORExt. Across the 123 responses to this item, QAs and QTs recommended four areas of improvement (in order of greatest frequency to least):

1. Option to electronically administer the test, so that scoring can be automatically computed
2. An assessment for students who cannot access reduced complexity academic tests due to severe limitations is needed, focusing on functional skills.
3. A new assessment (or new items) should be developed to better match the varying levels of abilities across the range of SWSCD, and
4. The math assessment should include more practical/life skills focused items (i.e., money, time) rather than complex mathematical concepts that are too advanced for this population (i.e., numerical coordinates, algebraic thinking)

The third qualitative response item asked respondents to describe their understanding and use of the curricular and instructional resources available through the lms.brtprojects.org website. Across the 123 responses, QAs and QTs provided two areas of descriptions (in order of greatest frequency to least):

1. Minimal knowledge and use of the resources due to various factors, for example, being unaware of its availability, not accessing the training, etc.
2. Provides resources to help teachers with test administrations (i.e., easyCBM) and write IEP goals and apply CCSS in their development.

Test Administration and Design

Overall, QAs and QTs overwhelmingly indicated that the OExt test was easy to administer and score, with about 97% agreeing or strongly agreeing with the quantitative survey statement. Similarly, respondents felt the test was accessible to their SWSCD, with about 77% of individuals agreeing or strongly agreeing, and that the performances required by the test items are appropriate behaviors to review for those students, with about 83% of individuals agreeing or strongly agreeing. Survey results suggest that the test is an efficient way to gather academic information for SWSCD.

Respondents' access to the necessary curricular to teach academic skills that are aligned to the Essentialized Assessment Frameworks appears to be limited, as most QTs and QAs indicated a mildly negative response with 53% of individuals either disagreeing or strongly disagreeing. This result is expected given that the curricular effort related to SWSCD is in its nascent period in Oregon. Efforts are underway to expand the availability and generalizability of these resources through coordinated efforts with ODE's curricular and content area specialists.

The majority of the respondents stated that they felt confident in interpreting the scores and the relative achievement level descriptors of the test, with 70% of individuals agreeing or strongly agreeing. Respondents' open-ended statements complemented these quantitative results, with over 80% of QAs and QTs commenting positively in some manner about that ease of test administration/scoring (75%) and item design/accessibility for SWSCD (61%). Furthermore, the majority of the responding QAs and QTs in Oregon felt the ORExt accurately reflected academic content that their SWSCD should be learning in school (72%).

As a higher percentage of positive responses is desired, we anticipate better alignment between assessment and instruction in the coming academic years, as new curricula are developed and teachers are further trained and more accustomed to the essentialized standards EsSt. Test items were written to align with EsSt designed to appropriately link to the Common Core State Standards (CCSS) and Next Generation Science Standards (NGSS), though reduced in depth, breadth and complexity. Overall, however, it appears that Oregon QAs and QTs are pleased with the ORExt test administration and design.

Educational Impact

Across all responding QAs and QTs, there was a general agreement regarding some of the potentially positive educational impacts from ORExt implementation. Respondents were positive that the ORExt increased educator understanding of and provided new assessment models from academic content. Additionally, many respondents believed that the ORExt improved the way in which classroom assessments are designed and administered, as well as increased the development of academic goals and IEP objectives

for their SWSCD. QA and QT opinions regarding other educational impacts of the OExt were not as positive. For example, respondents did not feel that OExt implementation positively impacted curricular and instructional approaches used for SWSCD or increased access to the general education curriculum. Through curricular approaches currently being designed and further training on test content and administration, positive opinion in these areas might grow in future academic years.

Discussion

Results from the fourth OExt consequential validity study point to historical concerns that are not possible to address, such as the ongoing tension between assessing life skills and academics, but also to some actionable steps with a focus toward continuous improvement. Respondents pointed to positive attributes of the OExt, especially those involving test administration and design and felt somewhat positive regarding various educational impacts of the OExt.

To better ensure the efficient administration and utility of the OExt, efforts should be made to effectively disseminate the knowledge and access of the curricular and instructional resources available through the lms.brtprojects.org website. The majority of the educators indicated that they were unaware of these resources, but voiced enthusiasm to obtain more information and training regarding their use and the steps to navigate the website.

Suggestions for Improvement

Regarding recommendations for improvement to the ORExt, QAs and QTs frequently cited the need to reassess math assessments in their appropriate levels of difficulty and functional applicability. Several educators indicated math questions should be less difficult with more practically focused items. This recommendation is further emphasized by those who specifically stated that items need to focus on functional skills of students who are still struggling with assessment content that has already been reduced in academic complexity, depth, and breadth. This is a concern that is consistent throughout AA-AAAS systems, but the ORExt is designed to assess academic content. While much academic content is indeed functional, a purely functional assessment would not meet the technical adequacy requirements of the Every Student Succeeds Act (ESSA, 2015). The item difficulty and person ability distributions for the mathematics assessment do not suggest that the assessments are too difficult, nor that they do not convey an appropriate range of functioning. This sentiment is likely a vestige of subjective experience that does not generalize to the wide, varied, population of SWSCD who participate in the ORExt.

Finally, QAs and QTs indicated that additional versions of the test should be developed to better match the ability level of those more severely impacted within the SWSCD population. Several educators have voiced that current items, despite the reduction in their complexity, depth and breadth, are not applicable for their students and continue to pose challenges in accurately assessing their level of performance. This concern has substantively been addressed with the development of the Oregon Observational Rating Assessment (ORora), which is an observational rating assessment for students whose ORExt testing is discontinued after they have met the minimum participation rule. In

addition, the range of item difficulties across each assessment conveys that the test is composed of a balanced number of low, medium, and high difficulty items that correspond well with the tested populations levels of ability.

Limitations

While the results from the 2017-18 Oregon Extended Consequential Validity Survey offer insights into the consequences of implementing the ORExt and point to actionable steps that can be taken to improve the assessment system, there are limitations to this study that affect the inferences that we can appropriately draw. First, the current results, as in years past, are subject to nonresponse bias because we cannot predict how those QAs and QTs who did not respond to the survey may have affected the results. A second limitation involves the nature of the four-point scale relative to the number of responding QAs and QTs—An even-value (four-point) rating scale used for the quantitative analyses did not allow respondents to remain neutral, a benefit given the survey was designed to "push" opinion positive/negative, but perhaps not refined enough to fully capture the full-range of respondents' thinking around the ORExt assessment. Overall, however, the consequential validity survey provided further evidence as to the impact of the ORExt and an additional basis for comparing the results to future years to define areas that need further improvement.

References

AERA, NCME, & APA. (1999). *The standards for educational and psychological testing*: American Educational Research Association.

Every Student Succeeds Act (2015). Public Law 114-95. Retrieved from https://www.congress.gov/bill/114th-congress/senate-bill/1177/text/pl

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement, 38*(4), 319-342.

Kleinert, H. L., Kennedy, S., & Kearns, J. F. (1999). The impact of alternate assessments: A statewide teacher survey. *The Journal of Special Education, 33*, 93-102.

Lane, S., Parke, C. S., & Stone, C. A. (1998). A framework for evaluating the consequences: Programs of assessment. *Educational Measurement: Issues and Practice, 17*(2), 24-28.

Messick, S. (1989). *Validity of test interpretation and use*. Retrieved from Princeton, NJ:

OESE. (2007). *Standards and assessments peer review guidance: Information and examples for meeting the requirements of the No Child Left Behind Act of 2001*. Washington, D.C.

Oregon Department of Education. (2014). *2013-2014 technical report: Oregon's alternate assessment system*. Retrieved from Salem, OR:

Popham, W. J. (1997). Consequential Validity: Right Concern - Wrong Concept. *Educational Measurement: Issues and Practice, 16*(2), 9-13. doi:10.1111/j.1745-3992.1997.tb00586.x

Roach, A. T., Elliott, S. N., & Berndt, S. (2007). Teacher perceptions and the consequential validity of an alternate assessment for students with significant cognitive disabilities. *Journal of Disability Policy Studies, 18*(3), 168-175.

Shepard, L. S. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice, 16*(2), 5-24.

Wilson, M. (2005). Construct maps *Constructing Measures: An Item Response Modeling Approach*. Mahwah, NJ: Lawrence Ehrlbaum Associates, Inc.

Table 1

Sex statistics for the 2017-2018 survey sample

Sex	*n*	*%*
Male	17	14
Female	103	84
Other (Gender Queer)	3	2
Total	123	100

Table 2

Age distribution for the 2017-2018survey sample

Age Range (Years)	*n*	*%*
20-25	3	1.7
26-30	12	10.2
31-35	11	9.3
36-40	12	10.2
41-45	22	17.8
46-50	15	11.9
51+	48	39
Total	123	100

Table 3

Educational experience for the 2017-2018 survey sample

Years Experience	*n*	*%*
0-3	11	9.2
4-7	19	15.8
8-11	8	6.7
12-15	21	16.7
16-20	27	20.8
21-25	15	12.5
26-30	13	10.8
31+	9	7.5
Total	123	100

Table 4

Regional representation of Oregon for the 2017-2018 survey sample

Region	*n*	*%*
Northeast (Pendleton, LaGrande Areas)	14	11.5
North I-5 Corridor (Portland, Beaverton, Hillsboro Areas)	44	36.1
North Coast (Astoria, Seaside, Lincoln City Areas)	3	2.5
Central (Bend, LaPine, Redmond, Madras, Prineville Areas)	8	6.6
Central I-5 Corridor (Eugene, Corvallis, Salem Areas)	39	32
Central Coast (Depoe Bay, Newport, Florence, Winchester Bay Areas)	2	1.6
Southeast (Burns Area)	2	1.6
South I-5 Corridor (Roseburg, Grants Pass, Medford Ashland Areas)	10	8.2
South Coast (Bandon, Port Orford, Brookings Areas)	0	0
Total	123	100

Table 5

Educational roles for the 2017-2018 survey sample

Years Experience	*n*	*%*
ORExt qualified assessor	96	78.2
ORExt qualified trainer	27	21.8
Total	123	100

Table 6

Quantitative and Qualitative Items for the 2017-2018ORExt Consequential Validity Survey

Survey item
1. The items in the Oregon Extended Assessment were easy for me to administer and score.
2. The items in the Oregon Extended Assessment were accessible for my students with significant cognitive disabilities (SWSCD).
3. The items in the Oregon Extended Assessment accurately reflect the academic content (what the student should know) that my students with significant cognitive disabilities should be learning, as defined by grade level content standards (CCSS/NGSS) and the Essentialized Assessment Frameworks.
4. The items in the Oregon Extended Assessment, which primarily ask students to match, identify, or recognize academic content, are appropriate behaviors to review to determine what my students with significant cognitive disabilities are able to do.
5. I have the curricula I need to teach academic skills that are aligned to the Essentialized Assessment Frameworks for my students who take the Oregon Extended Assessment.
6. I feel confident in interpreting the scores and their respective achievement level descriptors published for the Oregon Extended Assessments.
7. The implementation of the 2017-18 Oregon Extended Assessment has:
a. increased educator understanding of academic content for students with significant cognitive disabilities.
b. provided new models for assessing academics for students with significant cognitive disabilities.
c. positively impacted the curricular and instructional approaches used for students with significant cognitive disabilities in Oregon.
d. improved the manner in which classroom assessments are designed and implemented.
e. improved the learning outcomes for students with significant cognitive disabilities in Oregon.
f. increased access to the general education curriculum for students with significant cognitive disabilities.
g. increased the development of academic goals and objectives in IEPs for students with significant cognitive disabilities.

h. improved the alignment between IEP goals and objectives and state content standards and benchmarks.

15. *DESCRIPTIVE:* The following questions help us address test design concerns related to the Oregon Extended Assessment system as we invest in continuous improvement efforts.

16. Test administration for the Oregon Extended Assessment to me _____ (hours:minutes) on average for the following content areas this year. Note: do not count preparations of materials or data entry, only test administration.

17. Please describe what you appreciate most about the 2017-18 Oregon Extended Assessment.

18. Please recommend at least one improvement that could be made to the 2017-18 Oregon Extended Assessment.

19. Please describe your understanding and use of the curricular and instructional resources available through the lms.brtprojects.org website, Curricular and Instructional Materials for Students with Significant Cognitive Disabilities section.

Note. Six demographic items (21-26) were included at the end of the survey (displayed in Tables 1-6). Quantitative items (1-14) were positively-worded and used a four-point rating scale, where: 1 = strongly disagree, 2 = disagree, 3 = agree, and 4 = strongly agree. One quantitative item (16) required numerical input for response. Qualitative items (17-19) were open-ended responses. Items 15 and 20 were descriptive and served only to orient the respondent to the subsequent question block.