A Study of Practicality and Utility: Evaluations of the Online Research and Comprehension Assessments by both Practitioners and Policy Makers

PROJECT REPORT #11

Donald J. Leu, Elena E. Forzani, Cheryl Maykel, and Clint Kennedy
University of Connecticut

Portions of this material are based upon work supported by the U.S. Department of Education under Award No. R305G050154 and No. R305A090608. Opinions expressed herein are solely those of the authors and do not necessarily represent the position of the U.S. Department of Education.
In an online age of information, it is essential for all students to develop high-level reading and comprehension skills, including the ability required to conduct online research and write a short report (Common Core State Standards Initiative, 2012; Partnership for 21st Century Skills, 2006). As reading shifts from page to screen, and with online information use now such an important and ubiquitous part of our lives for learning and success (International Reading Association, 2009), increasing attention is being given to analyzing the nature of online reading, especially online research and comprehension (Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012; Hartman, Morsink, & Zheng, 2010; Zhang & Duke, 2008). Knowledge-based societies require citizens to be skilled in the effective use of online information for inquiry and communication (OECD, 2010; Rouet et al., 2009). Assessments are needed to measure the online research and comprehension skills required by students to succeed academically as well as for life-long learning (OECD, 2010.)

The ORCA Project (Leu, Kulikowich, Sedransk, & Coiro, 2009) developed 24 assessments in three formats and then evaluated the validity and reliability of each of the three formats and each of the 24 assessments in a pilot study among approximately 1,129 seventh grade students using representative samples of two states (Connecticut and Maine). From these results, a final set of assessments were selected and revised for psychometric testing during a validation year. These included two formats (multiple-choice, scenario based and simulation, performance based) and four different topics in science. In testing with 1,293 seventh grade students, representative samples from two states, these eight assessments demonstrated high levels of reliability, with KR-20 values ranging from .73 to .85 for the ORCA-Multiple Choice format and from .86 to .90 for the ORCA-Simulation format.

Both formats also demonstrated good estimates of validity although factor structures
were more complex for the ORCA-Closed than for the ORCA-Multiple Choice. CFI values ranged from .93 to .95 for the multiple-choice scales while RMSEA values were less than .05. For the ORCA-Closed scales, we also reported acceptable fit statistics in the ranges of .96 to 1.00 for CFI and with RMSEA equal to or less than .05.

Given the demonstration of acceptable levels of reliability and validity for the final set of eight ORCA assessments, an important question remained: To what extent are these assessments practical and useful to both practitioners and educational policy makers?

Without a demonstration of both practicality and utility, assessments such as those developed in this project are likely to be met with resistance in schools and, as a result, have limited potential to inform instruction. This issue is an important one since these assessments take place online and require the use of Internet computers, an important new context for schools to consider during assessment. They are especially important given early reports of frustration and difficulties with the online assessments being used for the CCSS assessments (Stephens, 2014; McLaughlin, Glaab, & Carrasco, 2014).

Four research questions were explored in this study. Two explore practicality:

1. To what extent are the assessments developed in the ORCA Project considered to be practical by educational practitioners?
2. To what extent are the assessments developed in the ORCA Project considered to be practical by educational policy makers?

Two others explore usefulness:

3. To what extent are the assessments developed in the ORCA Project considered to be useful by educational practitioners?
4. To what extent are the assessments developed in the ORCA Project considered to be useful by educational policy makers?

Method

Participants: Practitioners

Educational practitioners were included in this study to provide evaluations of practicality and usefulness based on the important realities of classroom contexts. Approximately 200 7th grade ELA and Science educational practitioners, from two Northeastern states and 100 different schools, including those who had participated in the pilot and validation year assessments, were invited to participate. Ninety-one teachers from this phase completed the survey, a response rate of 45.5%. Of those who designated a grade level for their position (82 out of 91), all indicated that they worked at the 7th and/or 8th grade level in their school. Eight-six of the participants listed their position as “teacher.” Five other participants listed their position as either: building science resource teacher, high school librarian, literacy coach, library media specialist, department liason library media or media teacher. The number of years that participants had been in their current position ranged from less than 1 year to 42 years, with a mean of 7.42 years. 48 participants taught in Connecticut, 40 in Maine, 2 in Massachusetts, and 1 in Texas. Twenty-one participants indicated that they worked in an urban school district, 45 indicated that they worked in a rural school district, 24 participants indicated they worked in a suburban school district, and one indicated the question was not applicable.

Participants: Policy makers
Educational policy makers were also included in this study to provide evaluations of practicality and usefulness based on their experiences with policy making at the school, district, regional, and state levels.

There were several phases to the recruitment of participants from each of the different levels of policy making. Approximately 40 middle school principals, whose schools had participated in the pilot and validation year assessments, from two Northeastern states, were invited to participate. 19 principals from this phase completed the survey, a response rate of 47.5%.

Next, we invited approximately 100 other principals from middle schools in each of these two states to participate. 12 responded for a response rate of 12% for this phase. Overall, our response rate was 22% for building principals.

We also recruited policy makers at other levels. Initially 40 members of the Governing Board and Research Partnership Teams who attend the meetings of the Governing Board of the New England and Island Regional Educational Lab were invited to participate. Thirty-nine members of this group completed the survey, a response rate of 97.5%.

Next, we invited each of these individuals to share the invitation with colleagues. This phase yielded 16 additional participants. It is not possible to calculate a precise response rate from this phase since we do not know how many individuals forwarded the email invitation.

Overall, a total of 86 policy makers participated in the survey from the school (n=35), district (n=25), or state levels (21). There was insufficient data to place the other five policy makers into one of these categories. Eighty worked in a New England state, two worked in New York, four worked in other parts of the country. Of those who reported the type of district context in which they worked, 12 indicated urban, 24 indicated rural, 25 indicated suburban. Ten indicated they had worked in education for ten years or less, 28 indicated they had worked in
eduction for 11-20 years, 25 indicated they had worked in education for 21-30 years, and 23 indicated they had worked in education for more than 30 years.

The Survey Instrument

The purpose of the survey was to evaluate the practicality and usefulness of the final assessments from the ORCA project in the eyes of key stakeholders: practitioners and policy makers. The survey was developed in two phases. In phase one, approximately 30 members of the Governing Board for the New England and Islands Regional Educational Lab (NEI-REL) were presented with a thorough description of the assessments from the ORCA project including a short experience with the assessments themselves. Members of this board include policy makers, many of whom are former practitioners, at local, regional, and state levels. Each member generated a ten-item list that they thought would best represent items to include in an evaluation of the practicality of the assessments among both practitioners and policy makers. Then, these lists were analyzed qualitatively, uning grounded theory (Corbin & Strauss, 2008) and analytic inductive analysis (Patton, 2005). This analysis showed that there were actually two main themes: practicality concerns and usefulness concerns. Practicality concerns appeared to cluster around several issues: 1) technology requirements; 2) scoring and reporting; and 3) administration time. Usefulness concerns also appeared to cluster around several issues: 1) the content of the assessments; and 2) the extent to which meaningful and relevant information is obtained.

In phase two, items generated from Governing Board members were revised for clarity and precision and included in each of the areas identified for practicality and usefulness above. A scheme for reporting the level for each item was also developed. Both aspects were then reviewed and revised by a panel of senior scholars in technology, measurement, and education.
The final survey consisted of 17 total items in three areas of practicality and three areas of usefulness. Each practicality item asked participants to respond on a 4-part scale:

- Not At All Practical
- Somewhat Practical
- Practical
- Very Practical

Each usefulness item asked participants to respond on a 4-part scale:

- Not At All Useful
- Somewhat Useful
- Useful
- Very Useful

Following each item, participants were asked to explain, in a single sentence, why they chose their response.

In addition, participants were asked to rank order the three areas of practicality in terms of importance. They were also asked to rank order the three areas of usefulness in terms of importance. The survey was delivered online and took approximately 40 minutes to complete.

The final survey structure was:

1. An introduction, including a video overview of the assessments.
2. Demographic information.
3. An Evaluation of the Practicality of the ORCAs (7 items)
   a. Technology requirements. (3 items).
   b. Scoring and reporting (2 items).
   c. Administration time (2 items).
4. An Evaluation of the Usefulness of the ORCAs (10 items)
   a. Content of the assessments/curriculum (3 items).
   b. The extent to which meaningful and relevant information is obtained. (5 items).
   c. The nature of the assessments - multiple choice or performance based (2 items).

5. Importance
   a. Ranked order, in terms of importance, of the three areas surveyed in relation to practicality (1 item).
   b. Ranked order, in terms of importance, of the three areas surveyed in relation to usefulness (1 item).

Analysis

For the purposes of this report, only scaled item data were analyzed and are presented here. Descriptive statistics (McMillan, 2014) were used to estimate both practicality and usefulness. For the criteria of practicality, results are presented in terms of the combined responses for “Practical” and “Very Practical.” Combined responses for “Practical” and “Very Practical” that totaled more than 50% of participants were considered to demonstrate practicality for each group of participants: practitioners and policy makers. For the criteria of usefulness, results are presented in terms of the combined responses for “Useful” and “Very Useful.” Combined responses for “Useful” and “Very Useful” that totaled more than 50% of participants were considered to demonstrate usefulness for each group of participants: practitioners and policy makers.
Results for the issue of importance permitted the ability to determine which, if any, dimension of either practicality or usefulness was most important to our two populations. If one dimension of either practicality or usefulness was more important than others, we wanted to know that and include it in our evaluations of practicality and usefulness.

Results

The complete results from the survey are presented in Table 1.
Table 1

*Results Of The Survey Of Practicality And Usefulness Of The ORCAs For Both Practitioners And Policy Makers.*

<table>
<thead>
<tr>
<th>Technology Requirements</th>
<th>Practitioners</th>
<th>Policy Makers</th>
</tr>
</thead>
<tbody>
<tr>
<td>5. Technology Requirements</td>
<td>% “Practical” or “Very Practical” Responses</td>
<td>% “Practical” or “Very Practical” Responses</td>
</tr>
<tr>
<td><strong>PRACTICALITY</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Technology Requirements</td>
<td>Practitioners</td>
<td>Policy Makers</td>
</tr>
<tr>
<td>6. These assessments require the use of individual laptops, desktops or Chromebooks for each student taking the assessment. In the school contexts most familiar to you, how practical is an assessment like this?</td>
<td>67.1%</td>
<td>70.9%</td>
</tr>
<tr>
<td>10. These assessments require adequate Internet access (wireless or Ethernet) that will permit 25-50 students to be online at the same time. No video downloads or streaming will be required. In the school contexts most familiar to you, how practical is an assessment like this?</td>
<td>83.5%</td>
<td>84.8%</td>
</tr>
<tr>
<td>12. In order to conduct the assessments, the following will need to be done:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. An adult will need to assign each student to an assessment online and obtain a code number for each student. This takes about 30 seconds per student.</td>
<td>81.4%</td>
<td>72.1%</td>
</tr>
<tr>
<td>b. Each student will need to enter his or her individual code number when prompted.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scoring and Reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. The <strong>ORCA-Closed</strong> assessment is performance-based. Students actually complete a research report, in an email or wiki, using online information. They are provided with a notepad and work with a search engine and text messaging. We assess performance with responses to 16 items. About half of the items are automatically scored by the program. About half are short, constructed response items and require hand scoring using a scoring rubric. In the school contexts most familiar to you, how practical is an assessment like this?</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Practitioners</strong></td>
<td><strong>Policy Makers</strong></td>
<td></td>
</tr>
<tr>
<td>% “Practical” or “Very Practical” Responses</td>
<td>% “Practical” or “Very Practical” Responses</td>
<td></td>
</tr>
<tr>
<td>52.8%</td>
<td>58.1%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administration Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>18. The <strong>ORCA-Closed</strong> is a performance-based assessment. Students actually complete a research report, in an email or wiki, using online information. They are provided with a notepad and work with a search engine and text messaging. It takes approximately 60 minutes to complete. In the school contexts most familiar to you, how practical is a 60-minute assessment like this?</td>
</tr>
<tr>
<td><strong>Practitioners</strong></td>
</tr>
<tr>
<td>% “Practical” or “Very Practical” Responses</td>
</tr>
<tr>
<td>50.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Administration Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. The <strong>ORCA-Multiple Choice</strong> uses a multiple choice assessment format with 16 items. Each item contains four multiple-choice responses. It takes approximately 20 minutes to complete. In the school contexts most familiar to you, how practical is a 20-minute assessment like this?</td>
</tr>
<tr>
<td><strong>Practitioners</strong></td>
</tr>
<tr>
<td>% “Practical” or “Very Practical” Responses</td>
</tr>
<tr>
<td>90.1%</td>
</tr>
</tbody>
</table>
USEFULNESS

Content of the Assessment/Curriculum
These assessments provide you with an evaluation of 7th grade students’ ability to conduct research on the Internet and write a short report. Students actually complete a research report, in an email or wiki, using online information. They are provided with a notepad and work with a search engine and text messaging. In the school contexts most familiar to you, how useful is it to have information on each student’s ability to:

22. Conduct research on the Internet and write a short report? 92.3% 95.4%

24. Conduct research on the Internet and write a short report in science? 90.1% 90.7%

26. Conduct the assessment above at the seventh grade level within the study of human body systems, as opposed to other domains in science? 72.6% 74.6%

The Extent To Which Meaningful And Relevant Information Is Obtained
28. The ORCA-Closed and ORCA-Multiple Choice assessments evaluate a student’s ability to conduct online research in science and write a short report in an email or a wiki. Performance is connected to several Common Core ELA anchor standards in reading and writing. In the school contexts most familiar to you, how useful is an assessment like this? 87.9% 89.4%

30. The ORCA-Closed and ORCA-Multiple Choice assessments evaluate online research skills and will provide a score report, with information on how each student performed, in each of four areas listed below. In the school contexts most familiar to you, how useful is an assessment that measures:
Useful or “Very Useful” Responses

<table>
<thead>
<tr>
<th>Activity</th>
<th>% “Useful” or “Very Useful” Responses</th>
<th>% “Useful” or “Very Useful” Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Reading to locate information using a search engine?</td>
<td>91.1%</td>
<td>94.2%</td>
</tr>
<tr>
<td>b. Reading to critically evaluate the reliability of information found on the Internet</td>
<td>93.4%</td>
<td>97.7%</td>
</tr>
<tr>
<td>c. Reading to synthesize information across multiple webpages</td>
<td>91.3%</td>
<td>97.6%</td>
</tr>
<tr>
<td>d. Writing to communicate a short report of research in an email or wiki</td>
<td>91.2%</td>
<td>96.5%</td>
</tr>
</tbody>
</table>

The Nature of the Assessment

32. The ORCA-Closed is a performance-based assessment where students actually complete a short research project online. They are evaluated during the research project on how well they perform various tasks. In the school contexts most familiar to you, how useful is an assessment like this?

<table>
<thead>
<tr>
<th>Importance</th>
<th>% “Most important” Responses</th>
<th>% “Most important” Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The technology requirements</td>
<td>42.9%</td>
<td>37.2%</td>
</tr>
<tr>
<td>b. The scoring and reporting</td>
<td>33.0%</td>
<td>31.4%</td>
</tr>
<tr>
<td>c. The administration time</td>
<td>24.2%</td>
<td>31.4%</td>
</tr>
</tbody>
</table>
38. You have already evaluated the **Usefulness** of the ORCAs in the three areas listed below. In the school contexts most familiar to you, how would you rank order these areas in terms of how important each is to a determination of practicality?

<table>
<thead>
<tr>
<th></th>
<th>Responses</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The content of the assessments /curriculum</td>
<td>37.4%</td>
<td>27.9%</td>
</tr>
<tr>
<td>b. The extent to which meaningful and relevant information is obtained</td>
<td>44.0%</td>
<td>64.0%</td>
</tr>
<tr>
<td>c. The nature of the assessments (performance based or multiple choice)</td>
<td>18.7%</td>
<td>8.1%</td>
</tr>
</tbody>
</table>

**Practitioners**

**The Evaluation of Practicality.** Three dimensions of practicality were evaluated in the survey by practitioners: technology requirements, scoring and reporting, and administration time. Each dimension was evaluated with several items that described the nature of the ORCA assessments. A majority of practitioners evaluated each descriptive item of the ORCAs in the three dimensions of practicality as either “Practical” or “Very Practical.” Two out of seven of the descriptive items had combined percentages for “Practical” and “Very Practical” responses of over 90%:

**Scoring**

- Item 16. Automatic scoring for the ORCA-Multiple Choice format (92.4%).

**Administration Time**

- Item 20. It takes approximately 20 minutes to complete the ORCA-Multiple Choice assessment (90.1%).
The two descriptive items of practicality with the lowest combined percentages of “Practical” and “Very Practical” responses still included a majority of practitioners:

*Scoring and Reporting*

Item 14. About half of ORCA-Closed assessment items are automatically scored and about half require hand scoring with a rubric (52.8%); and

*Administration Time*

Item 18. The ORCA-Closed assessment takes approximately 60 minutes to complete (50.6%).

**The Evaluation of Usefulness.** Three dimensions of the ORCAs’ usefulness were evaluated in the survey by practitioners: the content of the assessment/curriculum; the extent to which meaningful and relevant information is obtained; and the nature of the assessment. Each dimension was evaluated with several items that described the nature of the ORCA assessments. A majority of practitioners evaluated each descriptive item of the ORCAs in each dimension as either “Useful” or “Very Useful.” Six out of ten descriptive items in the dimension of usefulness had combined percentages for “Useful” and “Very Useful” responses of over 90%:

*Content of the Assessment/Curriculum*

Item 22. How useful is it to have information on each student’s ability to conduct research and write a short report? (92.3%);

Item 24. How useful is it to have information on each student’s ability to conduct research on the Internet and write a short report in science. (90.1%);

*The Extent To Which Meaningful and Relevant Information Is Obtained*

Item 30a. How useful is an assessment that measures reading to locate information using a search engine? (91.1%).
Item 30b. How useful is an assessment that measures reading to critically evaluate the reliability of information found on the Internet? (93.4%)

Item 30c. How useful is an assessment that measures reading to synthesize information across multiple webpages? (91.3%)

Item 30d. How useful is an assessment that measures writing to communicate a short report of research in an email or wiki? (91.2%)

The two descriptive items of usefulness with the lowest combined percentages of “Practical” and “Very Practical” responses still included a majority of practitioners:

**Content of the Assessment/Curriculum**

Item 26. Conduct the assessment above at the seventh grade level within the study of human body systems, as opposed to other domains in science? (72.6%); and

**The Nature of the Assessment**

Item 34. The ORCA-Multiple Choice is a multiple-choice assessment where students select one of four choices about a number of skills important to an online research project. In a school contexts most familiar to you, how useful is an assessment like this? (78.0%).

**The Evaluation of Importance.** Practitioners evaluated the importance for each of the three dimensions of practicality: the technology requirements, the scoring and reporting, and the administration time. The most important dimension of practicality for practitioners were the technology requirements. 42.9% of practitioners rated this dimension of practicality as most important. The next most important dimension for practitioners was the scoring and reporting aspects. 33.0% of practitioners rated this dimension of practicality as most important. The least
important dimension for practitioners was the administration time. Only 24.2% of practitioners rated this dimension of practicality as most important.

Practitioners also rated the importance for each of the three dimensions of usefulness: the content of the assessment/curriculum; the extent to which meaningful and relevant information is obtained; and the nature of the assessments (performance based or multiple choice). The most important dimension of usefulness for practitioners was the extent to which meaningful and relevant information is obtained. 44.0% of practitioners rated this dimension of usefulness as most important. The next most important dimension for practitioners was the content of the assessments/curriculum. 37.4% of practitioners rated this dimension of usefulness as most important. The least important dimension for practitioners was the nature of the assessments (performance based or multiple choice). Only 18.7% of practitioners rated this dimension of practicality as most important.

Policy Makers

The Evaluation of Practicality. Three dimensions of practicality were evaluated in the survey by policy makers: technology requirements, scoring and reporting, and administration time. Each dimension was evaluated with several items that described the nature of the ORCA assessments. With one exception, a majority of policy makers evaluated each descriptive item of the ORCAs in the three dimensions of practicality as “Practical” or “Very Practical.” The one exception was:

Administration Time

Item 18: The ORCA-Closed assessment takes approximately 60 minutes to complete (47.7%).
One out of seven descriptive items had combined response percentages for “Practical” and “Very Practical” of over 90%:

*Scoring*

Item 16. Automatic scoring for the ORCA-Multiple Choice format (92.4%)

**The Evaluation of Usefulness.** Three dimensions of the ORCAs’ usefulness were evaluated in the survey by policy makers: the content of the assessment/curriculum; the extent to which meaningful and relevant information is obtained; and the nature of the assessment. Each dimension was evaluated with several items that described the nature of the ORCA assessments. A majority of policy makers evaluated each descriptive item of the ORCAs in each dimension as either “Useful” or “Very Useful.” Seven out of ten descriptive items in the dimension of usefulness had combined response percentages for “Useful” and “Very Useful” of over 90%:

*Content of the Assessment/Curriculum*

Item 22. How useful is it to have information on each student’s ability to conduct research and write a short report? (95.4%);

Item 24. How useful is it to have information on each student’s ability to conduct research on the Internet and write a short report in science. (90.7%);

*The Extent To Which Meaningful and Relevant Information Is Obtained*

Item 30a. How useful is an assessment that measures reading to locate information using a search engine? (94.2%).

Item 30b. How useful is an assessment that measures reading to critically evaluate the reliability of information found on the Internet? (97.7%)

Item 30c. How useful is an assessment that measures reading to synthesize information across multiple webpages? (97.6%)
Item 30d. How useful is an assessment that measures writing to communicate a short report of research in an email or wiki? (95.6%)

The Nature of the Assessment

Item 32. How useful is an assessment like this? (90.7%)

The Evaluation of Importance. Policy makers evaluated the importance for each of the three dimensions of practicality: the technology requirements, the scoring and reporting, and the administration time. The most important dimension of practicality for policy makers were the technology requirements. 37.2% of policy makers rated this dimension of practicality as most important. The next most important dimension for policy makers were the scoring and reporting aspects. 31.4% of policy makers rated this dimension of practicality as most important. A similar percentage was reported for administration time. Only 32.4% of policy makers rated this dimension of practicality as most important.

Policy makers also rated the importance for each of the three dimensions of usefulness: the content of the assessment/curriculum; the extent to which meaningful and relevant information is obtained; and the nature of the assessments (performance based or multiple choice). The most important dimension of usefulness for policy makers was the extent to which meaningful and relevant information is obtained. 64.0% of policy makers rated this dimension of usefulness as most important. The next most important dimension for policy makers was the content of the assessments/curriculum. 27.9% of policy makers rated this dimension of usefulness as most important. The least important dimension for policy makers was the nature of the assessments (performance based or multiple choice). Only 8.1% of policy makers rated this dimension of practicality as most important.
Discussion

To What Extent Are The Assessments Developed In The ORCA Project Considered To Be Practical By Educational Practitioners?

The assessments were perceived by practitioners in this sample to be practical. Several patterns from the evidence suggest this conclusion. First, a majority of practitioners considered every descriptive item of the ORCAs (7 out of 7) in every dimension (three out of three) as either practical or very practical. Second, the dimension of practicality most important to practitioners were the “Technology Requirements.” More than two-thirds of practitioners considered the three items in this scale as either practical or very practical. The range was from 67.1% - 83.5%. Third, the item with the lowest percentage of practitioners who considered it practical or very practical (The ORCA closed assessment will take 60 minutes to complete.) was in the dimension of practicality least important to practitioners (Administration Time). The percentage of practitioners who still considered this item as either practical or very practical was greater than 50%. These results suggest that the assessments developed in the ORCA Project were perceived by practitioners as practical or very practical.

To What Extent Are The Assessments Developed In The ORCA Project Considered To Be Practical By Educational Policy Makers?

The assessments were also perceived by policy makers in this sample to be practical. Similar patterns from the evidence suggest this conclusion. First, with one exception, a majority of policy makers evaluated all of the descriptive items for the ORCAs (6 out of 7) as either practical or very practical. Second, the dimension of practicality most important to policy makers was also the “Technology Requirements.” Again, more than two-thirds of policy makers considered each of the three items in this scale as either practical or very practical. The range
was from 70.9% - 84.8%. Third, the item with the lowest percentage of policy makers who considered it practical or very practical (The ORCA closed assessment will take 60 minutes to complete) was in the dimension of practicality least important to policy makers (Administration Time). Nevertheless, the percentage of policymaker who considered this item as either practical or very practical (47.7%) approached a majority. All of these results suggest that the assessments developed in the ORCA Project were perceived by policy makers as either practical or very practical.

**To What Extent Are The Assessments Developed In The ORCA Project Considered To Be Useful By Educational Practitioners?**

The assessments were perceived by practitioners in this sample to be useful. Several patterns from the evidence suggest this conclusion. First, a majority of practitioners evaluated each descriptive item of the ORCAs in each dimension as either “Useful” or “Very Useful.” Second, six out of ten descriptive items in the dimension of usefulness had combined percentages for “Useful” and “Very Useful” responses of over 90%. Third, the dimension of usefulness most important to practitioners was the “The extent to which meaningful and relevant information is obtained.” With one exception, over 90% of practitioners considered the five items in this scale as either useful or very useful. The one exception had a combined response of useful or very useful that approached 90%. The range was from 87.9% to 93.4%. All of these results suggest that the assessments developed in the ORCA Project were perceived by practitioners as either useful or very useful.

**To What Extent Are The Assessments Developed In The ORCA Project Considered To Be Useful By Educational Policy Makers?**
Policy makers, too, found the assessments developed in the ORCA Project to be useful. Several patterns from the evidence suggest this conclusion. First, a majority of policy makers evaluated each descriptive item of the ORCAs in each dimension as either “Useful” or “Very Useful.” Second, seven out of ten descriptive items for usefulness had combined percentages for “Useful” and “Very Useful” responses of over 90%. Third, the dimension of usefulness most important to policy makers was the “The extent to which meaningful and relevant information is obtained.” With one exception, over 90% of policy makers considered the five items in this scale as either useful or very useful. The one exception had a combined response of useful or very useful that approached 90%. The range was from 89.4% to 97.6%. All of these results suggest that the assessments developed in the ORCA Project were perceived by policy makers as either useful or very useful.

Concluding Thoughts

Recent reports in the press (Stephens, 2014; McLaughlin, Glaab, & Carrasco, 2014) have indicated that the online delivered assessments developed for the CCSS initiative (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) have been met with frustration and difficulty by both practitioners and policy makers at local levels. There have been no systematic studies of these frustrations but reports of these frustrations are very common (See McLaughlin, Glaab, & Carrasco, 2014). It is not clear precisely what challenges appear to cause the greatest difficulty.

This study shows that the assessments of online research and comprehension developed in the ORCA Project appear to be perceived by both practitioners and policy makers in this sample as both practical and useful. The pattern of responses for practicality and usefulness are both clear and consistent. Only one descriptive item appeared where fewer than 50% of either
practitioners or policy makers considered the element either practical or very practical or useful and very useful. That item appeared in a dimension of practicality (administration time) for policy makers that was least important to them, among the three dimensions of practicality. In short, these assessments appear to have been developed in ways that will meet the practicality and usefulness needs of both practitioners and policy makers. Given this result, the design of these assessments may be useful to consider by others as literacy and literacy assessment shifts from page to screen.

References


**APPENDIX A**

**THE SURVEY OF PRACTICALITY AND USEFULNESS OF THE ASSESSMENTS DEVELOPED FOR THE ORCA PROJECT**

Evaluating the Practicality and Usefulness of the ORCAs (Online Research and Comprehension Assessments)

You are invited to participate in this survey to evaluate the practicality and usefulness of the ORCAs (Online Research and Comprehension Assessments).

Your participation in this study will require you to complete an online questionnaire. This survey will take you approximately 20-30 minutes to complete. Your participation will be anonymous and you will not be contacted again in the future. Your name will not be required, though if you wish to receive a $25 gift certificate to Amazon after completing the survey, you will be asked for your email address so we can send this to you.

This survey does not involve any risk to you. However, the benefits of your participation may impact society by providing useful information about the ORCA for schools.

You do not have to be in this study if you do not want to be. We will be happy to answer any questions you have about this study. If you have questions about this project or if you have a research-related problem, you may contact me at 860-486-0202 or via email at Donald.Leu@uconn.edu. If you have any questions about your rights as a research participant, you may contact the University of Connecticut Institutional Review Board (IRB) at 860-486-8802. The IRB is a group of people who review research studies to protect the rights and welfare of research participants.

If you agree to participate in this online survey, please click “Next” below. If you choose not to
participate, please close your browser window. Thank you!

INTRODUCTION:
The following video will give you an overview of the assessments. Watching the video will provide information that will be useful when responding to the survey questions. You may need to expand the video window to view the assessment portion of the video properly. There is no voice or sound on the video.

A description of video that appears on the survey.
Participants viewed a 10-minute video using QuickTime, a screen capture tool, that included several PowerPoint slides that introduced the assessments of online research and comprehension and provided a description of what is assessed. They also showed the performance-based nature of the Closed Assessment.

• Then, a summary of validity and reliability for the assessments was provided.
• Next, each assessment format (Closed, MC) was presented.
• Finally, the initialization procedures and the score sheet output was presented so reviewers could see how scores are reported and how many items will need to be hand scored in each format.

Purpose:
The purpose of this survey is to evaluate the practicality and usefulness of the ORCAs.

The ORCAs measure 7th graders’ ability to conduct online research in science, locate and evaluate online information, and write a short report in an email or wiki.

We want to know if you find these assessments practical and useful for schools.

EVALUATING ONLINE RESEARCH AND COMPREHENSION: A SURVEY OF PRACTICALITY AND USEFULNESS
[This will appear at the beginning of the survey, after the video.]
Demographic Information

Directions:
Please complete each question below, to the best of your ability, from the special vantage point that your position in education provides.

DEMOGRAPHIC INFORMATION

1. Please indicate your current role.
   - State-Level Policy Maker (State Commissioner, Member of the State Department of Education Office, or another State-Level Policy Shaper). Please specify your role in the text box to the right.
   - District-Level Policy Maker (District Superintendent or Member of the District Central Office). Please specify your role in the text box to the right.
   - Principal. Please indicate grade levels in the text box to the right:
     ______________________________________
   - Teacher. Please indicate grade level(s) and subject(s):
     ______________________________________
   - Other. Please indicate:
     ________________________________________

2. Please add any additional details about your response to Question 1 here. *(Text box)*

3. Please indicate the number of years you have worked in the type of role listed above.
   - *Drop down menu with choices: less than 1, 1, 2, 3.....45*

4. Please indicate the total number of years that you have worked in education.
   - *Drop down menu with choices: less than 1, 1, 2, 3.....45*

6. Please indicate the state in which you work. *(Drop down menu)*

7. Please indicate the school AND district in which you work. Type NA if you work at the state level. *(Text box)*

8. What is the type of district in which you work? Type NA if you work at the state level.
   *(Drop down menu with choices: Urban, Rural, Suburban)*

PRACTICALITY

Technology Requirements

9. These assessments require the use of individual laptops, desktops or Chromebooks for each student taking the assessment.
In the school contexts most familiar to you, how practical is an assessment like this?

☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
☐ Very Practical

10. Please explain, in a single sentence, why you choose the response you did in Item 8. *(Text box)*

11. These assessments require adequate Internet access (wireless or Ethernet) that will permit 25-50 students to be online at the same time. No video downloads or streaming will be required.

In the school contexts most familiar to you, how practical is an assessment like this?

☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
☐ Very Practical

12. Please explain, in a single sentence, why you choose the response you did in Item 10. *(Text box)*

13. In order to conduct the assessments, the following will need to be done:
   c. An adult will need to assign each student to an assessment online and obtain a code number for each student. This takes about 30 seconds per student.
   d. Each student will need to enter his or her individual code number when prompted.

In the school contexts most familiar to you, how practical is an assessment like this?

☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
☐ Very Practical

14. Please explain, in a single sentence, why you choose the response you did in Item 12. *(Text box)*

Scoring and Reporting

15. The *ORCA-Closed* assessment is performance-based. Students actually complete a research report, in an email or wiki, using online information. They are provided with a
notepad and work with a search engine and text messaging. We assess performance with responses to 16 items.

About half of the items are automatically scored by the program. About half are short, constructed response items and require hand scoring using a scoring rubric.

In the school contexts most familiar to you, how practical is an assessment like this?
☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
☐ Very Practical

16. Please explain, in a single sentence, why you choose the response you did in Item 14. (Text box)

17. The ORCA-Multiple Choice uses a multiple choice assessment format with 16 items. Each item contains four, multiple choice responses.

The program will automatically score all items.

In the school contexts most familiar to you, how practical is an assessment like this?
☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
☐ Very Practical

18. Please explain, in a single sentence, why you choose the response you did in Item 16. (Text box)

Administration Time

19. The ORCA-Closed is a performance-based assessment. Students actually complete a research report, in an email or wiki, using online information. They are provided with a notepad and work with a search engine and text messaging.

It takes approximately 60 minutes to complete.

In the school contexts most familiar to you, how practical is a 60-minute assessment like this?
☐ Not At All Practical
☐ Somewhat Practical
☐ Practical
20. Please explain, in a single sentence, why you choose the response you did in Item 18.
(Text box)

21. The **ORCA-Multiple Choice** uses a multiple choice assessment format with 16 items. Each item contains four multiple choice responses.

   It takes approximately 20 minutes to complete.

   In the school contexts most familiar to you, how practical is a 20-minute assessment like this?
   - [ ] Not At All Practical
   - [ ] Somewhat Practical
   - [ ] Practical
   - [ ] Very Practical

22. Please explain, in a single sentence, why you choose the response you did in Item 20.
(Text box)

---

**USEFULNESS**

**The Content of the Assessment/Curriculum**

These assessments provide you with an evaluation of 7th grade students’ ability to conduct research on the Internet and write a short report. Students actually complete a research report, in an email or wiki, using online information. They are provided with a notepad and work with a search engine and text messaging.

In the school contexts most familiar to you, how useful is it to have information on each student’s ability to:

<table>
<thead>
<tr>
<th></th>
<th>Not At All Useful</th>
<th>Somewhat Useful</th>
<th>Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. Conduct research on the Internet and write a short report?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. Please explain, in a single sentence, why you choose the response you did in Item 22.</td>
<td>(Text box)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. Conduct research on the Internet and write a short report <em>in</em></td>
<td>Not At All Useful</td>
<td>Somewhat Useful</td>
<td>Useful</td>
<td>Very Useful</td>
</tr>
</tbody>
</table>
25. Please explain, in a single sentence, why you choose the response you did in Item 24. (Text box)

26. Conduct the assessment above at the seventh grade level within the study of human body systems, as opposed to other domains in science? Not At All Useful Somewhat Useful Useful Very Useful

27. Please explain, in a single sentence, why you choose the response you did in Item 26. (Text box)

The Extent To Which Meaningful And Relevant Information Is Obtained

28. The ORCA-Closed and ORCA-Multiple Choice assessments evaluate a student’s ability to conduct online research in science and write a short report in an email or a wiki. Performance is connected to these Common Core ELA anchor standards in reading and writing:

<table>
<thead>
<tr>
<th>Reading Anchor Standards</th>
<th>Writing Anchor Standards</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Assess how point of view or purpose shapes the content and style of a text.</td>
<td>6. Use technology, including the Internet, to produce and publish writing and to interact and collaborate with others.</td>
</tr>
<tr>
<td>7. Integrate and evaluate content presented in diverse media and formats, including visually and quantitatively, as well as in words.</td>
<td>7. Conduct short as well as more sustained research projects based on focused questions, demonstrating understanding of the subject under investigation.</td>
</tr>
<tr>
<td>8. Delineate and evaluate the argument and specific claims in a text, including the validity of the reasoning as well as the relevance and sufficiency of the evidence.</td>
<td>8. Gather relevant information from multiple print and digital sources, assess the credibility and accuracy of each source, and integrate the information while avoiding plagiarism.</td>
</tr>
</tbody>
</table>
10. Read and comprehend complex literary and informational texts independently and proficiently.

In the school contexts most familiar to you, how useful is an assessment like this?

☐ Not At All Useful
☐ Somewhat Useful
☐ Useful
☐ Very Useful

29. Please explain, in a single sentence, why you choose the response you did in Item 28. (Text box)

30. The **ORCA-Closed** and **ORCA-Multiple Choice** assessments evaluate online research skills and will provide a score report, with information on how each student performed, in each of four areas listed in the table below.

In the school contexts most familiar to you, how useful is an assessment that measures these skills?

<table>
<thead>
<tr>
<th>Reading to locate information using a search engine</th>
<th>Not At All Useful</th>
<th>Somewhat Useful</th>
<th>Useful</th>
<th>Very Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading to critically evaluate the reliability of information found on the Internet</td>
<td>Not At All Useful</td>
<td>Somewhat Useful</td>
<td>Useful</td>
<td>Very Useful</td>
</tr>
<tr>
<td>Reading to synthesize information across multiple webpages</td>
<td>Not At All Useful</td>
<td>Somewhat Useful</td>
<td>Useful</td>
<td>Very Useful</td>
</tr>
<tr>
<td>Writing to communicate a short report of research in an email or wiki</td>
<td>Not At All Useful</td>
<td>Somewhat Useful</td>
<td>Useful</td>
<td>Very Useful</td>
</tr>
</tbody>
</table>

31. Please add additional details about your response to Item 30 here. (Text box)

The Nature of the Assessment
32. The ORCA-Closed is a performance-based assessment where students actually complete a short research project online. They are evaluated during the research project on how well they perform various tasks.

In the school contexts most familiar to you, how useful is an assessment like this?

☐ Not At All Useful
☐ Somewhat Useful
☐ Useful
☐ Very Useful

33. Please explain, in a single sentence, why you choose the response you did in Item 32. (Text box)

34. The ORCA-Multiple Choice is a multiple choice assessment where students select one of four choices about a number of skills important to an online research project.

In the school contexts most familiar to you, how useful is an assessment like this?

☐ Not At All Useful
☐ Somewhat Useful
☐ Useful
☐ Very Useful

35. Please explain, in a single sentence, why you choose the response you did in Item 34. (Text box)

36. You have already evaluated the Practicality of the ORCA assessments in the three areas listed below.

In the school contexts most familiar to you, please rank these areas in the order of how important each is to you in determining practicality.

1 = Most Important
2 = Next Most Important
3 = Least Important

☐ The technology requirements
☐ The scoring and reporting
☐ The administration time

IMPORTANCE
37. Please add additional details about your response to Item 36 here. (text box)

38. You have already evaluated the **Usefulness** of the ORCAs in the two areas listed below.

In the school contexts most familiar to you, how would you rank order these areas in terms of how important each is to a determination of practicality?

1 = Most Important  
2 = Next Most Important  
3 = Least Important

☐ The content of the assessments /curriculum  
☐ The extent to which meaningful and relevant information is obtained  
☐ The nature of the assessments (performance based or multiple choice)

39. Please explain, in a single sentence, why you choose the response you did in Item 38. (text box)

THANK YOU!

40. WE WILL SEND YOU A $25 GIFT CERTIFICATE TO AMAZON.COM IF YOU PROVIDE US WITH YOUR EMAIL ADDRESS BELOW. If you do not wish to provide your email address, please type ‘do not wish to provide’ in the text box. However, choosing this option means you will be unable to receive the gift certificate.

Email address: (text box)