

Fall 10-20-2010

Using Computer-based Testing with Young Children

Susan K. Barnes

James Madison University, barnessk@jmu.edu

Follow this and additional works at: https://opencommons.uconn.edu/nera_2010

 Part of the [Education Commons](#)

Recommended Citation

Barnes, Susan K., "Using Computer-based Testing with Young Children" (2010). *NERA Conference Proceedings 2010*. 22.
https://opencommons.uconn.edu/nera_2010/22

Abstract

The purpose of this study was to explore the feasibility of using computer-based tests (CBT) with children in preschool and kindergarten. Children were administered paper-and-pencil (PPT) and CBT versions of a rhyme awareness scale. After completing each assessment, each child shared individual reactions by selecting a card illustrating an emotion and participating in a brief interview. Parents and teachers completed short questionnaires describing each child's previous computer experience, fine motor skills, and ability to recognize and generate rhymes. An embedded mixed methods design was used to explore (a) to what extent children could complete the CBT independently, (b) how children reacted to the tests, and (c) how the results from the CBT and the PPT compared. Interview transcripts and field notes were used to more fully explain the test results. Findings indicated that young children needed help with the CBT. Children of all ages reported enjoying using the computer. Results provided preliminary evidence that test mode administration does matter. This CBT was more difficult than the PPT for all groups of children. These results have implications for test development and use. CBTs for preschoolers must be designed to meet their physical and cognitive developmental needs. Also, preschool children need adequate practice using computer hardware and software before they can reliably demonstrate their skills and abilities through CBT.

Keywords: Computer-based testing, early childhood assessments

Using Computer-based Testing with Young Children

In this era of increased accountability in education, there is a need for tools to use in assessing the abilities and instructional levels of young children. It is important to have reliable information describing children in order to make accurate decisions about referrals and appropriate learning environments for them. Computers have been used successfully to assess older children and adults, and there is much research comparing computer-based testing (CBT) to traditional paper-and-pencil testing (PPT) with older students and adults (e.g., Pomplun, Frey, & Becker, 2000; Wang, Jia, Young, Brooks, & Olson, 2007). Because CBT has been effectively used in secondary schools and higher education to efficiently test students for many years (Becker, 2000), it is tempting to simply assume that CBT could be used with children under eight years of age with equal success. However, unlike the abundance of CBT research done with older or special needs students, there is a dearth of available research focusing on the issues of computer-based testing with typically developing young children (Barnes, 2010). Thus it is important to gather information regarding the feasibility of using CBT with this population. Of specific importance are: the degree to which young children are able to complete the assessments with minimum individual support from adults, the children's reactions to CBT, and the usefulness of the results. These questions need to be researched within the context of the constraints of what we know about testing young children in general.

Purpose of the Study

This paper reviews components of the author's doctoral dissertation research (Barnes, 2010). The purpose of the study was to investigate the feasibility of using computer-based tests (CBT) with typically developing young children. The potential benefits of CBT are promising, but first, educators who use assessment results to plan instruction for young children need to

know if the investment of funds, time, and effort to develop computer-based assessments will yield information that is valid, meaningful and useful to them. This study explored factors that impact the merit of CBTs for young children, specifically, the level of adult supervision that children between the ages of three and six need when taking a CBT, young children's reactions to CBT, and the comparability of their CBT and PPT scores.

Research Questions

To address this purpose, the following research questions were investigated:

1. To what degree are typically developing young children able to complete the computer-based test version of the rhyme awareness subscale of the Preschool Phonological Awareness Literacy Screening (Invernizzi, Sullivan, Meier, & Swank, 2004) (CBT-PALS) independently? For those children who are not able to complete the CBT-PALS independently, how much and what kinds of support do they need in order to finish the test?
2. How do children react to the rhyme awareness subscale of the CBT-PALS?
3. Do young children demonstrate the same level of rhyme awareness when tested using the paper and pencil version of the Preschool Phonological Awareness Literacy Screening (PPT-PALS) as they do on the CBT-PALS? How are CBT-PALS results related to the children's gender and age?

Method

Participants

Exactly 100 children participated in the study. Their ages ranged from two years and eleven months to six years and two months. They all live in the Shenandoah Valley. The demographics of the children in the public kindergartens were different from the city child care centers. The children in preschools were one or two years younger than the children in the

kindergartens. The city child care centers have a larger proportion of children from minority groups, including English language learners. Targeting these diverse classrooms resulted in a sample of participants who represented the population of interest. Table 1 contains descriptions of the participating schools.

Table 1

Description of Participants Recruited for the Study

	Number of classrooms participating	Number of students in participating classrooms	Percent of students eligible for free or reduced lunch	Percent of students non-white	National Center for Education Statistics locale type
A Elementary	6	99	32%	11%	Rural: Fringe
B Elementary	2	45	36%	4%	Rural: Fringe
C Elementary	5	84	20%	10%	Rural: Fringe
D Elementary	3	49	42%	5%	Rural: Town
E Elementary	1	15	30%	3%	Rural: Fringe
F Preschool	2	25	n/a	5%	Small City
G Preschool	3	36	n/a	20%	Small City
Total	26	456			

See Table 2 for the number of children from each age group and gender from the individual schools. Not every child agreed to take both the PPT and the CBT. Data from all participants who completed the CBT were used to explore children's ability to complete the CBT independently and children's reactions to the assessments. However, only data from children who completed both assessments were included in the component of the study that addressed how the scores from the CBT and PPT compared.

Table 2

Description of Participants

	Number of classrooms participating	Number of boy participants	Number of girl participants	Number of children from this school
A Kindergarten	5	6	7	13
B Kindergarten	2	9	11	20
C Kindergarten	3	8	8	16
D Kindergarten	2	4	6	10
E Kindergarten	1	5	2	7
F Preschool	2	6	7	13
G Preschool	3	10	11	21
Total	18	48	52	100

Instruments

The following section describes the instruments and procedures used for data collection, including the type of data yielded by each one. Table 3 includes an overview of this information, with each instrument presented in the approximate order in which it was used for data collection. Also provided is a description of how each source of data contributed to the study.

Table 3

Contributions of Individual Data Sources

Instrument	Data Type	Contribution
Parent Questionnaire	Quantitative – classifies students into groups Quantitative – ranks students by experience and age	Provided data for test mode effect analysis (age, gender, computer experience, input device use)
PPT-PALS Rhyme Scores	Quantitative	Provided data for test mode effect analysis
CBT-PALS Rhyme Scores	Quantitative	Provides data for test mode effect analysis
Observation Field Notes	Qualitative	Described children’s actions and language while completing the assessments
Observation Checklist	Quantitative	Provided counts of behaviors children exhibited while completing the assessments
Debriefing Session Interview Notes	Qualitative	Captured the children’s language explaining their reactions to the assessments
Emotion Cards	Quantitative	Allowed children to share their own reactions to the assessment, without using words, by selecting a picture
Teacher Surveys	Quantitative	Identified other factors that may have explained test mode effect

Procedures

The research questions were addressed using a mixed methods approach, simultaneously using assessment methods familiar to early childhood professionals and considered to be “best practices” as described by the National Association for the Education of Young Children (NAEYC, 2009), while exploring the use of a new computer-based test. Data on children’s level of independence, reaction to the testing, and level of achievement were measured using direct observation, sub-scale scores for rhyme awareness from the Phonological Awareness Literacy

Screening (PALS) for Preschoolers (Invernizzi, Sullivan, Meier, & Swank, 2004), CBT scores from the same PALS sub-scale, and information provided by parents and teachers.

Figure 1 provides an overview of the research design. When one type of data set provides a supportive role because the primary data type is not sufficient to address the research questions in

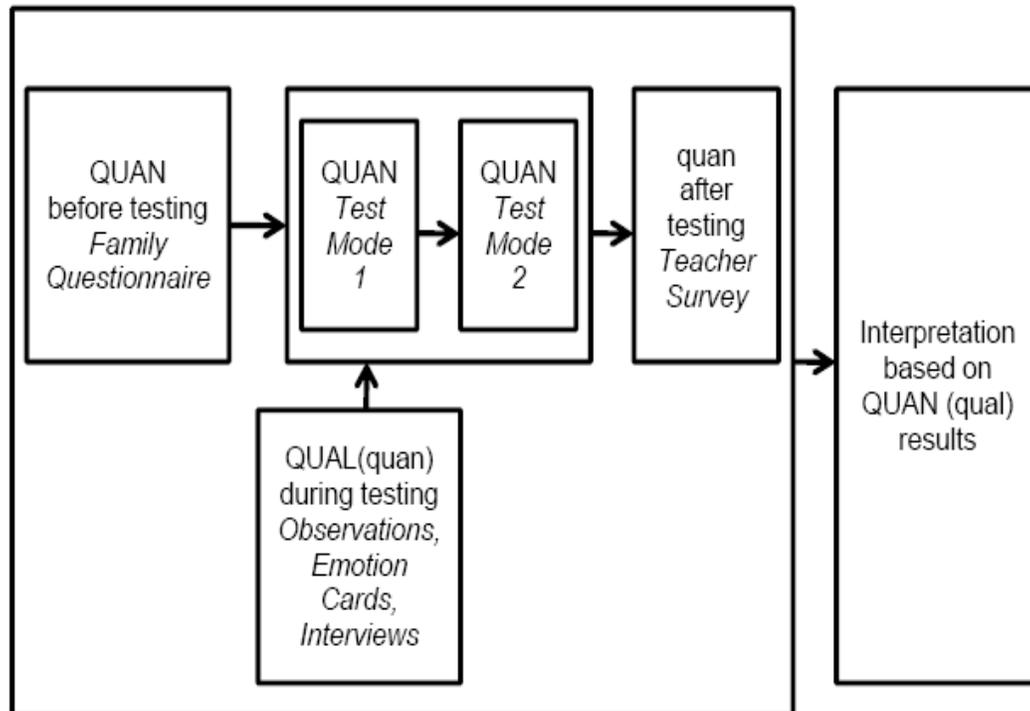


Figure 1. The embedded design of the mixed methods model used to examine the feasibility of CBT with young children.

the study, this can be referred to as an embedded design (Creswell & Plano Clark, 2007). Some data collection activities in the study occurred simultaneously and some sequentially. This visual guideline for the mixed method study uses the notation system used extensively in the mixed method literature wherein arrows indicate sequence, upper case letters indicate the method with the greater emphasis, lower case letters indicate the secondary method, and parentheses surrounding a

method indicate that the parenthetical method is supporting the other method (Creswell & Plano Clark, 2007). Test Mode 1 and Test Mode 2 refer to the CBT and PPT. The order of test administration was randomly assigned to counterbalance the order effect.

Can Young Children Test Independently?

This question addressed the degree to which young children could complete a CBT on their own, and what kinds of support they required. Do they have the attention span, confidence and competence to complete a CBT without depending on the individual attention and reassurance they typically get from their teachers in assessment situations? Children typically use computers for learning activities and entertainment. Unlike an assessment, these activities do not require them to persist until the end of the task. The level of persistence children exhibit, a trait closely related to attention span, varies widely from child to child (Kagan & Snidman, 2004). Typically, when children abandon a computer game for some other activity, they experience little, if any, encouragement to return to the computer activity. This study examined the degree to which children persist in the assessment activity when provided verbal prompts by the computer program.

In the booklet version of the test, children were seated at the table with the researcher. In this administration, if children became distracted, the examiner redirected their attention back to the test. While the language of the scripted prompts was the same for each test format, the CBT lacked some of the benefit of the proximity of the teacher to the student. The mere nearness of a teacher helps students to attend to the tasks at hand (Flicker & Hoffman, 2006). The CBT had the same set of motivating statements embedded in the test. Unlike the teachers, the CBT system was not able to discriminate between children who were gazing at the screen and children who are distracted from the assessment. In each format, regardless of the accuracy of their responses or the length of time between responses, children receive the same neutral acknowledgements and

statements to encourage them to continue with the test.

To measure the degree to which young children persist and complete the CBT independently without individual support from an examiner, the researcher formally observed the children taking the CBT. The researcher recorded the type and frequency of support children needed to stay on task using prepared field note forms and checklists. The researcher administered the paper tests and noted the same kinds of support the children needed while taking the test in that mode. The CBT and the PPT versions of the test were administered in the same setting, an area familiar to the children either in or near their classrooms.

How Do Children React to the CBT?

The second question addressed how children react to the CBT. To answer this question each child was briefly interviewed immediately after each test. During these debriefing sessions, open-ended questions allowed children to respond using their own words. Then each child was asked to select a card from a series of cards that had pictures of faces expressing emotions ranging from anger to joy. Finally, each was invited to share any explanations about the card he or she chose or other comments.

How Do Test Results Compare?

An important aim of this study was to determine if CBT-PALS scores for young children were comparable to the PPT-PALS scores. In this study, children's rhyme awareness was measured using two different methods, the administration of the Pre-kindergarten PALS subscale (PreK PALS) in paper format and as a CBT. To describe how CBT-PALS scores relate to other variables such as age and gender, information provided by parents and teachers regarding demographics and computer experience was examined.

Limitations

This study is limited in that it focused on measurements in the language domain of child development. One cannot generalize the results of this study to other domains, such as mathematical thinking. In fact, the results are limited to this particular phonemic awareness assessment. Another limitation is related to the sample of convenience. While the children in this study represent diversity in economic status, rural and urban residence, and levels of computer experience, they are all from the same general geographic region of the country, the Shenandoah Valley of Virginia. While they all speak English, it is a second language for some of the children. The data for the study were nested, with children nested within classrooms, classrooms within schools and schools within districts. However models for examining nested data were not appropriate for a sample this small. Finally, the software used to administer the CBT presented limitations in the functionality of the test.

Contributions to Research and Practice

Given the increased demand for evidence of student achievement and the high cost of individually administered tests, information regarding the efficiency, reliability and validity of scores from CBT is extremely valuable. Unlike the computer assessments used with older students and special needs children, there is an inadequate body of research addressing issues related to assessing typically developing young children using computers.

Results from this study provide evidence to address these questions surrounding the feasibility of using CBT with young children. This information is important because assessments such as the PALS are used by many public schools to identify children needing additional screening and school-based services. Early identification of children who need services is important. Delayed identification of children needing additional services can lead to the delayed

implementation of interventions for this population (National Joint Committee on Learning Disabilities, 2006). There is an extensive body of research documenting that early intervention results in a reduction in special education spending and a decrease in the number of children repeating grades (Barnett, 1985; Belfield, 2004; Conyers, Reynolds, & Ou, 2003). Studies have also found substantial long-term savings in terms of lower welfare costs and lower juvenile justice costs (Campbell & Rainey, 1994; Diefendorf & Goode, 2005). While timely identification is important, educators must take care to select assessment methods that provide accurate results. Errors in identification of children can result in their being placed in learning environments that are inappropriate because they do not meet the children's needs. In summary, educators need multiple methods to assess the knowledge and skills of young children. Many traditionally used assessment approaches are extremely time-consuming and administered in inconsistent ways. CBT has the potential to provide an unbiased and efficient source of information to those who plan learning experiences for young children.

Results

Can Young Children Test Independently?

Children expressed the kind of help they needed by asking questions, making comments, and by exhibiting behaviors that indicated that they were unable to proceed on their own. The children needed two major kinds of support: help with using the mouse and with knowing what to do next to proceed through the CBT. Kindergarteners were more independent than the preschoolers during the assessments. While some preschoolers were able to complete the CBT with little to no assistance, many of their peers needed support from an adult in order to finish. Girls asked more questions than boys. Figure 2 provides an overview of the kinds of support preschoolers and kindergarteners needed in order to complete the CBT.

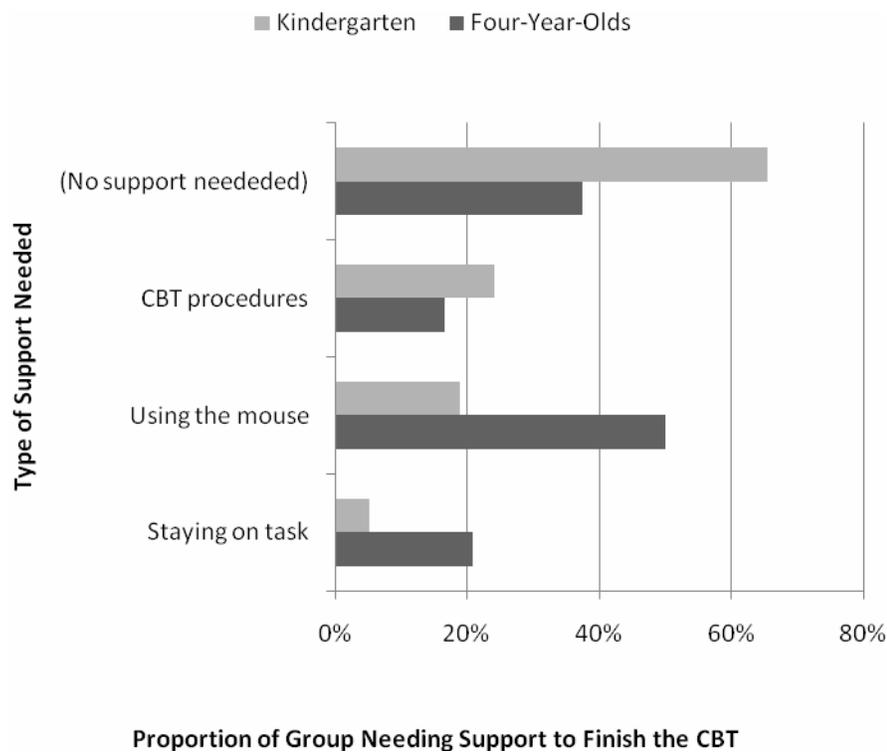


Figure 2. Proportion of kindergarteners and four-year-olds who needed support to finish the CBT. N=58 kindergarteners. N= 24 preschoolers.

How Do Children React to the CBT?

Overall, the children said they liked the CBT. Some explained that they liked using the computer and doing the rhyming activities. An overwhelming majority of children selected the highest-ranking positive emotion card, however, most children were not able to explain why they selected the card with the joyful face. When using the mixed methods integrated approach and considering the authentic language of the children along with the card selection data, the child’s perspective on the CBT experience becomes less positive than the quantitative card selection data may initially indicate. Several children compared the two tests and remarked that they preferred

the PPT because it was quicker. The children’s reactions and the comments the children made during the debriefing sessions were analyzed and several themes emerged. Figure 3 illustrates the card selection data within each of the verbal reaction themes.

How Do Test Results Compare?

The third question examined how the test results compared across administration modes. The scores on the CBT were consistently lower than the scores on the PPT for all age groups. Item analysis revealed that nine of the 10 items on the CBT had a higher difficulty index than when

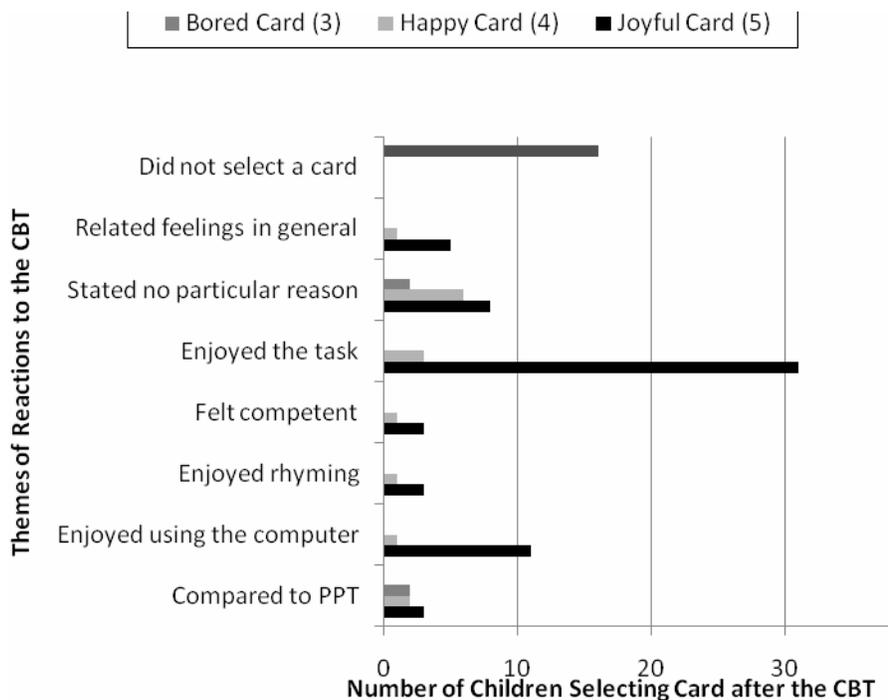


Figure 3. Number of children selecting emotion cards and expressing language within the most common CBT themes.

presented in the PPT format. A practice effect, a rise in the second test score, was expected because children took the tests in the same day and there was an opportunity for learning to take place during the first assessment.

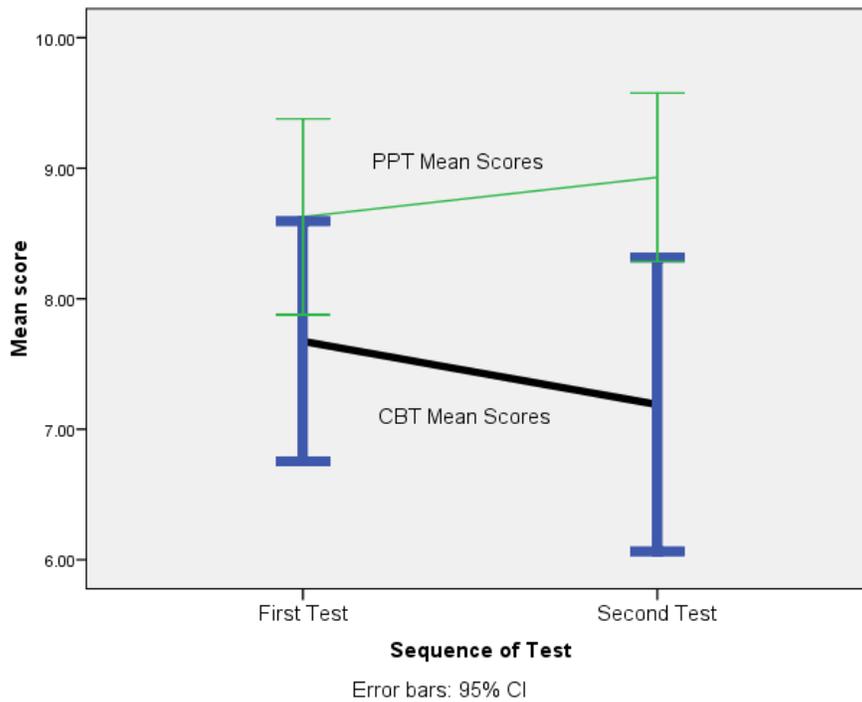


Figure 4. The relationship of mean total test score for PPT and CBT with the order of test administration. N=86.

Test administration order appeared to have an interaction with the mode of administration. See Figure 4. When the CBT followed the PPT, CBT scores were lower than when the CBT was administered first. When the PPT followed the CBT, PPT scores were higher than when the PPT was first. The mean score on the PPT for students who took the CBT first was slightly higher than the mean score for students who took the PPT as their first test. However, the results were just the opposite for mean CBT scores. The mean score for the CBT when it was administered as the second test was nearly a full point lower than the mean score for the CBT when it was administered first. These results indicate that the children were either fatigued or bored with the CBT process by the time the second test was administered.

The Kruskal-Wallis test results show a statistically significant difference in the CBT scores for preschoolers and kindergarteners. No difference was found between the CBT and PPT scores for kindergarteners or between the PPT scores for preschoolers and kindergarteners. Differences across the gender groups were small and not statistically significant. See Table 4.

Table 4
Kruskal-Wallis Test for Differences on Test Scores

Group 1	<i>n</i>	Mean rank	Group 2	<i>n</i>	Mean rank	Chi-square	df	Asymp. sig.
Male	41	41.49	Female	45	45.33	.62	1	.429
CBT first	43	39.21	PPT first	43	47.79	3.12	1	.077
Preschool CBT	12	9.71	Kindergarten CBT	31	26.76	17.88	1	.000
Preschool PPT	16	17.50	Kindergarten PPT	27	24.67	4.72	1	.030
Preschool CBT	12	9.96	Preschool PPT	16	17.91	6.68	1	.010
Kindergarten CBT	31	27.68	Kindergarten PPT	27	31.59	1.16	1	.281

Note: new Bonferroni adjusted alpha= 0.006

Discussion

Many traditionally used assessment approaches are extremely time-consuming and administered in inconsistent ways, and CBT has the potential to provide an unbiased and efficient source of information to those who plan learning experiences for young children. The benefits of standardized computer-based testing, such as quick access to results and the objectivity of scores, make the use of CBT tempting to administrators and policymakers who wish to have results that are easy to compare across populations. For busy teachers, the potential to gather information on the children’s progress without spending valuable one-on-one time with

individual children administering a test is appealing. But, no matter how efficient the administration or how objective the results are, if scores do not accurately represent what the children actually know or what they can in fact do, then these measures are of little value to the educators who use assessment information to design instruction for children. This study identified some important issues to consider when using CBT with young children.

First, in order for teachers to benefit from the efficiency of an automated assessment that does not require a lot of teacher support, children need to be able to progress through the assessment independently. In this study, the preschool children struggled with using the mouse. Simpler input devices, such as touch screens or voice recognition systems, would promote more independence. Children should also have ample opportunities to practice using the devices before testing. Kindergarten children, who generally did not have problems with the input device, needed many reminders to stay on task. These children wanted to explore the CBT to see what functions were available. As they learned the limited use of the program, their interest waned. In this study, a deliberate attempt to make the CBT as similar as possible to the PPT precluded the use of colorful graphics or animation. Including these innovations could make the CBT more engaging and help to maintain the children's interest.

While the children in this study responded positively to the CBT when asked to select an emotion card to signify how they felt about the assessment, their remarks about the length of the test and their preference for the quicker PPT may indicate that the CBT was not as enjoyable as their card selections indicated. Also, when comparing the test results, scores on the CBT were consistently lower than the PPT, regardless of the order in which the tests were administered. Possibly, the CBT was measuring something other than rhyme awareness, such as boredom or

fatigue, for preschoolers. This supposition is supported by the fact that CBT test scores for the preschoolers and kindergarteners were statistically different when the PPT scores were not.

Summary

This study examined the feasibility of using CBT with children who have not yet started the first grade. Kindergarten children were able to perform the assessment nearly as well on the CBT as they did on the traditional paper version of the test. However, most children under four years of age were not able to complete a CBT independently. The children reacted positively to the CBT and seemed to enjoy using the computer, but the test results indicate that the CBT was more difficult and may be measuring something more than rhyme awareness for the preschoolers. Research into the level of effort that children expend while taking the CBT should be examined further. Questions that remain are: (a) is taking the CBT more mentally and physically taxing than the PPT, and (b) does the level of effort needed to complete the CBT impact scores in a consistent or predictable way. These concerns need to be clearly investigated before test developers and users can have confidence that the use of CBT is appropriate for young children.

References

- Barnett, W. S. (1985). Benefit-cost analysis of the Perry Preschool program and its policy implications. *Educational Evaluation and Policy Analysis*, 7, 33-42.
- Barnes, S. K. (2010). *Using computer-based testing with young children*. (Ph.D. dissertation). Retrieved from ProQuest Dissertations and Theses database. (Publication No. AAT 3403029.)
- Becker, H. J. (2000). Who's wired and who's not: children's access to and use of computer technology. *The Future of Children: Children and Computer Technology*, 10, 3-31.
- Belfield, C. R. (2004). *Early childhood education: How important are the cost-savings to the school system?* (Report for Center for Early Care and Education) Retrieved October 8, 2009 from <http://www.plan4preschool.org/documents/ny-cost-saving.pdf>
- Campbell, A., Ramey, C. (1994). Effects of early intervention on intellectual and academic achievement: A follow-up study of children from low-income families [Electronic version]. *Child Development*, 65, 684-698.
- Conyers, L. M., Reynolds, A. J., & Ou, S. (2003). The effect of early childhood intervention and subsequent special education services: Findings from the Chicago Child-Parent Centers. *Educational Evaluation & Policy Analysis*, 25(1), 75-95.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.
- Diefendorf, M., & Goode, S. (2005). *The long term economic benefits of high quality early childhood intervention programs*. Retrieved November 9, 2008 from RAND Corporation at http://www.rand.org/pubs/research_briefs

- Flicker, E., & Hoffman, A. (2006). *Guiding children's behavior: Developmental discipline in the classroom*. New York, NY: Teachers College Press.
- Invernizzi, M., Sullivan, A., Meier, J., & Swank, L. (2004). *Phonological awareness literacy screening: PreK*. Charlottesville, VA: University of Virginia.
- Kagan, J., & Snidman, N. (2004). *The long shadow of temperament*. Cambridge, MA: Belknap Press of Harvard University Press.
- NAEYC. (2009). Joint position statement from the National Association for the Education of Young Children and the National Association of Early Childhood Specialists in State Departments of Education: Where we stand on curriculum, assessments and program evaluation. Retrieved September 25, 2010 from <http://www.naeyc.org/files/naeyc/file/positions/StandCurrAss.pdf>
- National Joint Committee on Learning Disabilities. (2006). Fact sheet. Retrieved November 8, 2008 from <http://www.ldonline.org>
- Pomplun, M., Frey, S., & Becker, D. (2000). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension [Electronic version]. *Educational and Psychological Measurement, 62*, 337-353.
- Wang, S., Jiao, H., Young, M. J., Brooks, T. E., & Olson, J. (2007). A meta-analysis of testing mode effects in grade k-12 mathematics tests [Electronic version]. *Educational and Psychological Measurement, 67*, 219-238.