J·T·L·A

The Journal of Technology, Learning, and Assessment

Volume 4, Number 2 · November 2005

Knowing What All Students Know: Procedures for Developing Universal Design for Assessment

Leanne R. Ketterlin-Geller

www.jtla.org

A publication of the Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College

J·T·L·A
Volume 4, Number 2

Knowing What All Students Know: Procedures for Developing Universal Design for Assessment

Leanne R. Ketterlin-Geller

Editor: Michael Russell
russelmh@bc.edu
Technology and Assessment Study Collaborative
Lynch School of Education, Boston College
Chestnut Hill, MA 02467

Copy Editor: Kevon R. Tucker-Seeley
Design: Thomas Hoffmann
Layout: Aimee Levy

JTLA is a free on-line journal, published by the Technology and Assessment Study Collaborative, Caroline A. & Peter S. Lynch School of Education, Boston College.

Preferred citation:

Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *Journal of Technology, Learning, and Assessment, 4*(2). Available from http://www.jtla.org

Abstract:

Universal design for assessment (UDA) is intended to increase participation of students with disabilities and English-language learners in general education assessments by addressing student needs through customized testing platforms. Computer-based testing provides an optimal format for creating individually-tailored tests. However, although a theoretical basis for universal design is well established, little practical information is available to assist test developers in creating and implementing universally designed tests. This article discusses the application of universal design to assessment and describes how these principles are applied to a test of 3rd grade mathematics ability. I present the steps involved in conceptualizing, constructing, and implementing a universally designed test in anticipation that test developers, state department assessment coordinators, and other researchers will benefit from this application. Recommendations for future research and development efforts to create accessible computer-based learning environments for all students are explored.

J·T·L·A

Knowing What All Students Know: Procedures for Developing Universal Design for Assessment

Leanne R. Ketterlin-Geller
University of Oregon

Introduction

Testing requires a confluence of skills, some that represent the construct of interest and others that reflect ancillary processes involved in engagement with the material. For example, multiple-choice mathematics tests require reading skills to access the information in order to solve the problems. Therefore, the construct being tested (mathematics) becomes intertwined with access skills not intended for measurement (reading) resulting in possible construct-irrelevant variance in student scores (Messick, 1989). Students with deficiencies in access skills may be unable to demonstrate their knowledge and abilities in the tested domain.

One approach to resolving this situation is to build appropriate tests at the onset by applying principles of universal design to assessments and providing the users with customized tests based on individual needs. Using a multi-method approach for assessing access skills, Universally Designed Assessment (UDA) systems determine students' needs and match those needs with appropriate testing conditions. These systems allow students to demonstrate their abilities in the tested domain unencumbered by problematic access skills.

Accommodations support students with limited access skills by changing the way the test is given or taken. At present, accommodations are applied to assessment tasks subsequent to conceptualization and development of the test for the general education population. Several problems limit the effectiveness of these externally imposed accommodations. Some problems are attributable to the external nature of these test accommodations, such as their insensitivity to individual differences or the packaging of multiple accommodations by test publishers that may not be necessary or even beneficial. Others are caused by inconsistent assignment and delivery procedures that limit the effectiveness of accommodations. It follows that the educational community needs a systematic mechanism that will improve the process for identifying, implementing, and monitoring the use of accommodations on assessment tasks. UDA is a test development system that has the potential to address these issues.

UDA is based on the premise that environmental features combine with personal characteristics to influence the success of user interactions. Through careful planning and consideration of the wide range of student characteristics, UDA seeks to amend the environment by creating individually tailored tests based on individual needs. The customized testing environment systematically embeds format changes within the assessment tool and anchors these changes to student performance by providing a multi-method design to accurately assess basic skills necessary for test completion. Consideration of the variety of student characteristics throughout the conceptualization, construction, and implementation stages of test development supports the needs of the greatest proportion of learners and reduces the need for external accommodations (Tindal & Crawford, 2003). Thus, UDA provides direct access to content information and student performance data without retrofitting existing materials or relying on imprecise decision-making methods. As such, UDA captures the knowledge and abilities of the widest range of students with the least amount of external accommodations or modifications to the test structure or format.

Ironically, no universal definition for UDA has been established in the literature. The different conceptualizations range from broadly stated principles that result in fixed test formats (Thompson, Johnstone, & Thurlow, 2002) to adaptable testing environments that provide individualized supports to address the needs of the range of student characteristics (Dolan, Hall, Banerjee, Chun, & Strangman, 2005). In this paper, I suggest that UDA is best defined as an integrated system with a broad spectrum of possible supports so as to provide the best environment in which to capture student knowledge and skills. Much like researchers investigating Universal Design for Learning (Rose, Meyer, Rappolt, & Strangman, 2002), this notion of universal design considers both the physical and cognitive environments. As such, this view of UDA allows for greater flexibility in the design of assessment tasks to support the wide range of student needs within the general education setting.

Computer-based technology presents an efficient tool for customizing assessments to meet individual needs within a universally designed environment. Computer-based tests administer items to the learner through an interactive computer environment. Item format can vary from static representation of the problem to different multimedia presentations, such as video or animation. Responses are captured directly by the computer in a variety of selection and production response formats. These variations from the standard paper-and-pencil format increase the accessibility of the information and allow greater flexibility for the learner (Dolan, 2000). Efficiency of scoring and implementation decreases administrative costs and reduces possible sources of error (Burk, 1998). Additionally, revisions and alternate forms can be made easily with minimal expense (Thompson, Thurlow, & Moore, 2002). Advanced computer applications, such as computer-adaptive tests (CAT), may also be used to increase the precision and efficiency of the test (Rudner, 1998). Although not required for creating a universally designed test, these qualities of computer-based technology make it an ideal mechanism for measuring student performance in a universally designed environment.

In this paper I discuss how features of the environment influence the ways in which people interact with places and events. I explore principles of universal design applied to educational testing environments as a mechanism to enhance the presentation, accessibility, and usability of assessment tools for all students and promote an inclusive learning environment. I present an example case in which the principles of universal design are applied to a 3rd grade mathematics test. I consider procedural descriptions as well as test specifications that may act as guidelines for test developers, state department assessment coordinators, and other researchers interested in developing and implementing universally designed tests.

People and the Environment: The Changing Nature of Interactions

The structure and format of an environment influences the quality and quantity of user interactions. Whether referring to the entrance of a building, a computer interface, or a test of mathematics ability, environmental features can hinder or enhance the user's experience, thereby limiting or extending access to this forum. Contributing to these experiences is the interplay between environmental elements and individual characteristics. Elements of the environment include features such as the directions for the user, layout of the information or material, or physical operation of the equipment. Personal attributes, such as cognitive abilities, physiological and physical characteristics, or individual preferences, also influence the scope of environmental interactions (Story, Mueller, & Mace, 1998). Conflict between environment and user limits the accessibility of information or materials. In essence, every environment requires specific user characteristics for successful entrance, much like a password. Without the proper features, the user is denied access. For example, if written material (environmental element) is presented to a person with a severe visual impairment (specific individual characteristic), access is denied. Furthermore, until the environment is changed, the information remains inaccessible.

Fluid characteristics and expectations of the user interact with the structure and format of the environment. For example, a person with cystic fibrosis may need an elevator to reach the second floor of a building on a "bad" day but may use the stairs on a day when symptoms are minimal. Thus, an environment without an elevator is accessible depending on the state of the individual. Similarly, a student learning to read may only gain understanding from text written using simple language until decoding skills are mastered. Material containing complex language during this stage of development may limit the student's access to the information. Thus, the nature of the environment may prohibit successful engagement by the user.

Considering that many users with both fixed and fluid characteristics interact with any given environment, the variety of individual specifications is limitless. Therefore, flexibility in structure and format is essential to meet the diversity of individual needs. Universally designed environments are intended to accomplish this goal by focusing on changing the environmental interface, often through the use of computer-based technology to meet the range of user needs. Before applying these principles to assessment, the theoretical underpinnings of universal design must be discussed.

Universal Design: Enhancing Interactions Between People and the Environment

Universal design is a conceptual extension of historical mandates calling for equal access to environments for people with disabilities (Story, Mueller, & Mace, 1998). The Americans with Disabilities Act of 1990 significantly impacted the lives of individuals with disabilities by requiring equal access to all public goods and services and prohibiting disability-related discrimination in employment (U.S. Department of Justice, 2003). Retrofitting existing buildings increased access for individuals with disabilities but at the expense of costly, cumbersome designs (Pisha & Coyne, 2001). Although functional and legally acceptable, these specially constructed features separated the populations based on physical need (Higbee, 2001). Considering these suboptimal results, Ronald Mace, an architect with a physical disability, proposed integrating high-access features into buildings from the beginning (Thompson, Johnstone, &

Thurlow, 2002). Incorporating structural and format features that serve the greatest percentage of the population during the design process provides all users with seamless access to environments while mediating the high cost of redesigning current structures and the social demoralization of segregation based on physical need (Mace, Hardie, & Place, 1991). These designs are classified as "universal" (Center for Universal Design, 1997).

Characteristics and expectations of all members of the population drive the structure and format of a universally designed environment, thereby ensuring that individual needs are satisfied without compromising the experiences of others. The Center for Universal Design (CUD) (1997) recommends designing environments that are equitable, flexible, and intuitive in their use by individuals regardless of their physical approach mechanisms. For example, automatic entrance doors provide people with physical disabilities greater ease when entering or exiting buildings. At the same time, people with strollers, carts, canes, or even an arm load of groceries benefit from this feature. Not only does the automatic door allow for greater access for people with disabilities, it also assists other users. In this way, changes to an environment may support the needs of multiple users. Additionally, the CUD (1997) states that environments should provide clear and explicit information, accept imprecise nonessential movements or actions, and incorporate maximally negotiable spaces that minimize unnecessary physical or cognitive efforts. What results is "[t]he design of products and environments [that are] usable by all people, to the greatest extent possible, without the need for adaptation or specialized design" (Center for Universal Design, 1997, paragraph 3).

Extending the reference of "products and environments" to educational settings pushes the application of universal design beyond the physical classroom and school-based environments and into cognitive learning spaces (Rose, Meyer, Rappolt, & Strangman, 2002). This added complexity requires an expanded view of the principles of universal design as proposed by CUD to recognize learning processes. As such, not only must the wide range of physical characteristics be accounted for in the design and development of educational settings but so must the range of cognitive and sensory abilities. Ultimately, the intention of universal design in education is to provide an equitable learning environment that is accessible by all students to the greatest extent possible.

Although the primary goal of universally designed environments is to permit equal access by all members of the population, a truly universal situation is not always possible (Story, Mueller, & Mace, 1998). Unique personal circumstances may preclude some individuals from interacting with the current condition of some environments. In these situations, additional modifications to the structure and/or format may be necessary.

For instance, a mathematics test may be appropriately created to address the needs of students with low reading abilities and students with visual and auditory impairments. However, a student with autism may need additional support such as individual administration. Thus, the term "universal" must be interpreted as an intended, yet not fully attainable, goal of universal design.

Comparable Performance Across All Students

Despite the limited utility of external accommodations due to variations in their delivery and imprecise assignment procedures, they continue to be a mechanism for supporting students with disabilities and English-language learners in general education assessments. The inherent confusion surrounding the use of these test changes compromises the validity of the interpretations made from test results. In contrast, applying the principles of universal design to tests of academic achievement is the "best way to increase participation in general state and district assessments" (Thompson & Thurlow, 2002, p. 2) of students with disabilities and English-language learners through the use of embedded accommodations. Because of the standardization of accommodation assignment and administration procedures, scores obtained from individuals in these populations can be combined with scores of students receiving general education services. Aggregating test scores in this manner can lead to meaningful interpretations of student performance. It follows that this process should increase the validity of the subsequent educational decisions resulting from observed scores.

External Accommodations: Can We Get Comparable Performance?

Currently, students with diverse needs are provided with external testing accommodations designed to remove physical, cognitive, or sensory barriers that may inhibit understanding or expression of domain-specific knowledge (National Center on Educational Outcomes, 2003). Acting as gatekeepers, these personal attributes may prohibit successful interactions with the information or instructional task. For example, students with limited reading proficiencies encounter significant difficulties when interacting with multiple-choice mathematics tests (Helwig, Rozek-Tedesco, Tindal, Heath, & Almond, 1999). Accommodations support the learner by mediating the effects of inhibitory characteristics without altering the construct under investigation. By changing the medium through which information is presented, allowing alternate response formats, altering the external environment, or adjusting the timing of the testing situation (Hopper, 2001), accommodations allow students direct access to the targeted domain (Elliott, Kratochwill, & McKevitt, 2001). With well-suited accommodations, the accuracy of the assessment results increase, leading to more appropriate inferences about student ability. Although consistent with the theories supporting UDA, the imprecise nature of the delivery mechanism and potential misassignment of externally imposed accommodations may compromise student success and jeopardize the integrity of the decision-making system.

To date, test accommodations are made by retrofitting existing assessments. Teachers, publishers, and test developers typically use accommodations with previously created materials designed for the general education population. Although this process may remove some barriers to accessing the information, student performance may suffer as a result of inappropriate changes to the testing materials (Fuchs, Fuchs, Eaton, Hamlett, Binkley, & Crouch, 2000). For accommodations to be beneficial, format changes must be specific to the individual's needs (Helwig & Tindal, 2003). For example, a student with a visual impairment may benefit from information presented in large text or Braille; however, this same student may be distracted or confused by auditory presentation of information. An individualized approach to selecting accommodations may significantly increase student access to information, resulting in more accurate inferences about student knowledge and abilities.

An Individualized Educational Program (IEP) team makes decisions that govern the use of external accommodations for an individual student (Lowe & Reynolds, 2000). Composed of members of the educational community who are familiar with the student, an IEP team evaluates the student's characteristics in order to create, implement, and monitor instructional goals (Hickman, 2000). Input from a variety of sources – including parent preference, teacher's prior experience, and inferences about student performance – is considered when determining the use of external accommodations for achieving these benchmarks (Fuchs & Fuchs, 1999). Consistent and reliable procedures, however, currently are not in place to assist the team in making these recommendations (Shriner & Destefano, 2003). This lack of uniformity limits the trustworthiness and reliability of accommodation decisions. It follows that IEP team decision-making should be mediated by the use of standardized procedures that adequately measure a student's need for an accommodation. A systematic process of identifying appropriate accommodations may provide a better way to meet students' individual needs, and thus provide greater opportunity for all students to have equal access to the tested content.

Embedded Accommodations: Getting Comparable Performance

Universally designed tests support the needs of the larger population of learners by considering student characteristics in the conceptualization, creation, and implementation stages of test development (Thompson, Johnstone, & Thurlow, 2002). As such, the intersection between personal characteristics and environmental features are thoroughly evaluated to design the most appropriate testing interface. Although universally designed assessments support many students with limited access skills, one test cannot appropriately measure all students. “Because there is an inevitable interaction between the representational demands of the medium and individual capacities of the students, for each student there will be some inadvertent effect of the medium in which the assessment occurs” (Dolan, 2000, p. 47).

In these situations, UDA systems provide a flexible testing platform to identify and deliver items that are customized to fit the individual’s needs. This flexibility within the testing system sets UDA apart from externally imposed accommodations by embedding accommodations within the test for seamless implementation and assignment of appropriate accommodations based on the individual’s current needs. Assignment of accommodations is determined during the (pre-) testing process. Input from a variety of sources such as teacher, parent, and student surveys as well as student performance on a series of basic skills tests administered at the start of the testing session can be evaluated for deficiencies in the requisite access skills needed to succeed in the tested construct. Used in this manner, universally designed assessments provide a data-based mechanism for identifying and then delivering appropriate accommodations at the individual level.

Flexibility in Testing: Creating a Universally Designed Test

To highlight the application of universal design principles to test development, we created a universally designed test of 3rd grade mathematics ability. What follows is a description of the steps involved in conceptualizing, constructing, and implementing a UDA. Research on the usability and accessibility of this test is presented. However, findings from validity studies can be found in Ketterlin-Geller, 2003. Many of the steps discussed below mirror those followed during the creation of any test; however, the primary difference that sets UDA apart from other test development procedures is the conscious and deliberate consideration of individual needs within the design of the testing environment.

In brief, the first step in creating any test is to clearly define the construct under investigation. Items need to be written in a format that is appropriate for the construct under investigation and should include minimal influence from extraneous, non-construct-related variables. Certain formats lend themselves to efficient measurement of specific domains, but are not appropriate for others. For instance, multiple-choice tests increase accessibility by reducing the complexity of user-interface interactions (Bennett, 2002; Burk, 1998). However, such items may pose limitations to some students or may jeopardize inferences about student ability in performance-based domains such as writing. To address these issues and the range of student characteristics, when appropriate, test developers may consider embedding flexibility within the item format so that users could select the response mode that is most appropriate to their needs. As this demonstrates, a priori consideration of the item format is essential for accurate measurement of student knowledge and abilities within a given domain.

To address the influence of environmental features, universally designed tests should provide clear and explicit information as well as a testing environment that is equitable, flexible, and intuitive (Thompson, Johnstone, & Thurlow, 2002). The test-development team needs to consider the most efficient testing platform to deliver the item and test specification. If using a computer-based testing platform, items can be delivered in a fixed item order using computer-based technology or through a flexible system using computer-adapted technologies. Although computer-adapted tests (CAT) may increase efficiency and accuracy (Rudner, 1998), students are not able to return to the previous item once they have advanced. As this may pose a problem for some students, computer-based tests may be more appropriate for specific populations.

Additional design and measurement issues should be discussed so as to permit the widest range of options without changing the tested construct. General supports should be embedded within the test to support all students and specific accommodations need to be identified that will address the needs of students with limited access skills.

Next, the test should be reviewed and evidence for the validity of the uses of the test should be evaluated. Items need to be field tested with members of the target population to determine appropriate item functioning and to identify sources of bias that may interfere with student performance. Statistical analyses such as differential item functioning (DIF) may highlight items that contain sources of cultural or disability-related bias that unfairly influences performance. Item response theory (IRT) modeling provides a mechanism for examining DIF by comparing the probabilities of students with different group identifications and known ability levels getting the same item correct (Hambleton, Swaminathan, & Rogers, 1991). These procedures ensure that the items accurately measure the intended construct thereby generating meaningful data for decision making.

Finally, members of the educational community should provide input about the appropriateness, ease of use, and feasibility of the assessment system before the results are used to make decisions. Feedback from students as well as parents, teachers, administrators, and representatives of child advocacy groups can come in the form of bias reviews, pilot testing, focus groups, or surveys and may lead to valuable insights into the functioning of the test for members of the target population.

What follows is a description of how we applied these procedures during the conceptualization, construction, and implementation stages of a universally designed test. I present an example case of a computer-based universally designed 3rd grade mathematics test that included a multi-method system of accommodation assignment. The reading-based accommodations included simplified language, read-aloud, and read-aloud with simplified language. Once assigned, accommodations were seamlessly integrated into the delivery of the test. By focusing on clearly defining the construct, conceptualizing and constructing a testing platform as well as item and test formats to reflect the needs of a diverse population, and soliciting and evaluating feedback from the target population and members of the community, we created a flexible testing platform that provided customized support based on individual needs.

Definition of the Construct

To begin this project, we clearly defined the construct of the test as well as the intended uses of the results. Although the UDA system used in this project incorporated additional basic skills tests, the purpose of the assessment system was to measure student ability in mathematics. The target domain was defined as the knowledge and skills identified by the state content standards for 3rd grade: calculations and estimations, algebraic relations, statistics and probability, geometry, and measurement. The test included items that sampled a wide range of student ability. The target population for the UDA included students receiving special education services as well as those students in the general education system; diverse cognitive abilities were anticipated. With these specifications articulated, we began conceptualizing the most appropriate format for delivery.

Test Conceptualization and Construction

To efficiently measure student proficiency in mathematics while addressing the needs of the diverse student population, the test platform as well as the format of the items and the test were carefully considered. A team of researchers, psychometricians, and computer programmers evaluated the specifications to create a test that reflected the principles of universal design.

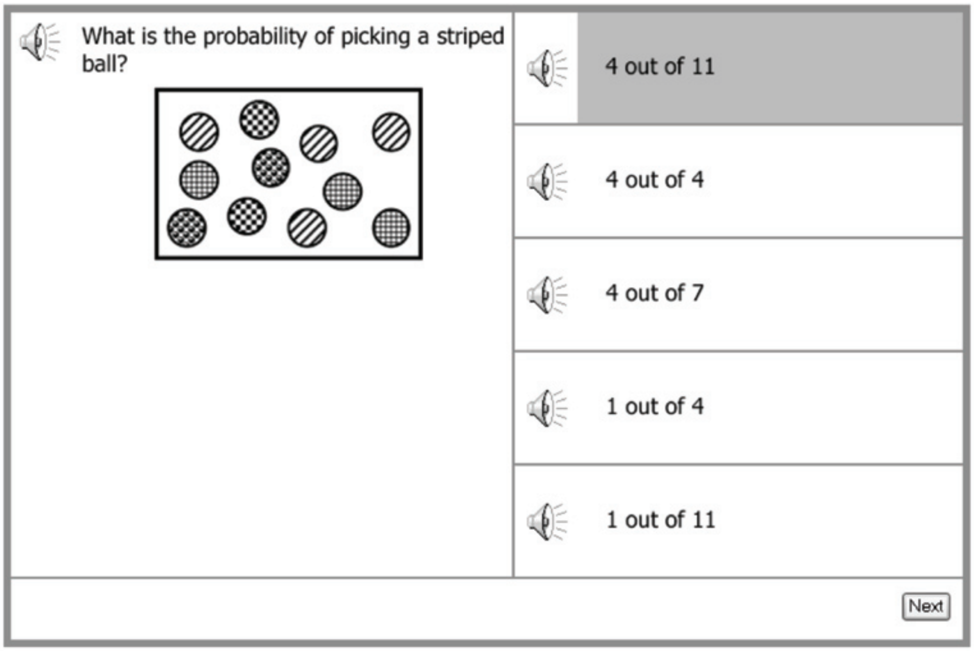
Testing Platform

To deliver the 3rd grade mathematics test with maximum versatility within one testing environment, we designed an efficient testing platform that accommodated for the test specifications in a manner that increased the accessibility for all users. Although computer-based technology is not necessary for creating a universally designed environment, we selected a computer-based format in order to create an equitable and flexible environment for use by all participants. Software-design experts collaborated with assessment experts to review the purpose of the test and devise technological solutions for maintaining the goals while providing accessibility and flexibility. We embedded support mechanisms such as a practice items and easy to navigate pages for students with limited computer skills to avoid introducing unnecessary sources of error. Feasibility issues such as equipment costs, technical support, and planning needs were evaluated and addressed in order to measure the intended construct without jeopardizing student access. What resulted was a testing platform that all students could easily interact with thereby increasing the accessibility of the test.

Item Format

To measure mathematics proficiency for the universally designed test, we considered both the item type as well as the presentation of the item in light of the characteristics of the target population. For consistency with the state assessment format and to reduce the complexity of the testing environment (Bennett, 2002; Burk, 1998), test items were written using a selection-response, multiple-choice format. We incorporated five answer choices for each item to reduce the measurement error caused by guessing. Problems were formatted with the question on the left of the screen and the answer choices listed vertically on the right. Figure 1 illustrates this configuration. An answer choice was activated when the student tabbed to or clicked anywhere within the box containing the answer. Allowing the selection of responses via the computer mouse or keys on the keyboard extended the range of motor skills and physical abilities needed to access the material (Thompson, Thurlow, & Moore, 2002). Color and luminance changes indicated the selection of an answer. The students were allowed to change their answer choices at any time prior to submitting their responses. Responses were not recorded until the student progressed to the next screen, thereby allowing the student to review his/her selection prior to submitting the answer.

Figure 1: Interface Design for the Computer-formatted Tests



Although items could be delivered using standard computer-based technology, CAT technology was selected to maximize efficiency and increase the accuracy of ability estimates (Rudner, 1998). The effects of errors on student ability estimates were minimized due to the precision of the item selection procedure as specified by the computer-adapted algorithm. An estimate of ability was calculated after each response and used to select a new item. The resulting response pattern was used to deliver the next most appropriate item associated with the student's performance. Thus, item-based errors were well accommodated in the universally designed mathematics test.

Test Format

Along with item format, we applied the principles of universal design to the presentation of items, sequence of tasks, and overall structure of the test to create a test that accommodated the range of student characteristics. Access to the information was available auditorily as well as in written text to provide redundancy of material (Center for Universal Design, 1997) as well as to provide supplemental support to users with low reading skills.

All directions and prompts were written in simplified text (see Tindal, Anderson, Helwig, Miller, & Glasgow, 2000 for simplification procedures). To enhance the visual presentation for young learners and learners with diverse needs, the screen design accommodated all relevant information in black 18-point sans serif font, without requiring the user to move the focal point or scroll. A 5-item test of mouse-maneuvering capabilities was administered prior to testing to provide practice navigating through the test on the computer. The range of computer skills tested mirrored those needed to interact with the test, thereby mediating the effects of differential experience with computers (American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME), 1999). Students who needed additional practice using the mouse were directed to retake the mouse-maneuvering test and/or received additional assistance from the administrators. These features made the content of the test accessible for students with varying physical needs within the same testing situation, thereby avoiding segregation based on physical abilities.

Embedded Accommodations

For those students who had difficulties accessing the information or expressing their knowledge and abilities, supplemental accommodations were embedded into the UDA system. For the universally designed mathematics test, additional accommodations were available to students with poor word-recognition and decoding skills and/or poor reading-comprehension skills. Each student took a series of basic skills tests to assess his or her abilities in reading. Students with low word-recognition and decoding skills were administered the mathematics items in simplified text, thereby reducing the level of English language proficiency needed to interpret the problems (Stansfield, 2002). Students with low reading-comprehension skills were presented with the mathematics items in an auditory as well as text format to promote access to the material through listening comprehension (Helwig, Rozek-Tedesco, & Tindal, 2002). These items were formatted with an icon that resembled a speaker to the left of the text. The student activated the read-aloud option by clicking on the speaker icon. No limit was placed on the number of times a student could listen to the directions or problems. These design features attenuated individual characteristics that may negatively impact a user's ability to access the tested content, thereby increasing the perceptibility of the mathematics problems.

Test Verification and Revision

Before actually using the universally designed test to make educational decisions, we gathered input from a variety of sources to evaluate the adequacy and appropriateness of the items and format. We collected construct-, content-, and criterion-related evidences for validity to ensure that the measures were meaningful and reliable. Three grade-level teachers as well as three content experts analyzed each item and concluded that a majority of the items were appropriate in their content, language, and format. Inappropriate items were excluded or rewritten prior to testing. All items were pilot tested with approximately 500 students from diverse backgrounds and with varying needs. Internal consistency reliability data indicated that the items were measuring the same construct. Results were analyzed using a 2-parameter Item Response Theory (IRT) model. Each item was examined for fit with the model, quality of the distracters, and sensitivity. Items that diverged from the model with a significance level less than $\alpha = .10$ were excluded. Additionally, items with poorly written distracters were removed because they may inappropriately confuse a test taker and therefore not adequately reflect a student's mathematics ability. Finally, the sensitivity of the items was assessed in order to identify items that provided additional information for distinguishing among ability levels. As a result of these analyses, inappropriate items were removed (for detailed procedures see Ketterlin-Geller, 2003).

Additionally, we solicited input about the appropriateness of the computer-testing interfaces by incorporating members of the educational community throughout the test development process. Administrators, teachers, parents, child advocacy group members, and students participated in focus groups where they interacted with the test and provided direct feedback on the ease and flexibility of the computer interface. Participants responded to specific questions about the functioning of the testing system as well as the appropriateness of the test features for the diverse range of student needs. We used results to amend the existing item format and computer-based testing platform (see Ketterlin-Geller, Alonzo, & Tindal, 2004 for detailed results). By incorporating feedback from a variety of constituents in the development process, the universally designed math test reflected the needs of the target population to the greatest extent possible.

In summary, this example case illustrates how we applied the principles of universal design to test development procedures to create a testing environment that was suitable for a majority of the population. Using computer-based technologies, we created a UDA that provided a flexible testing platform that was amendable to individual characteristics. Students requiring additional support received accommodations that were embedded in the testing platform and specifically targeted to their needs. What resulted was a more equitable testing environment designed to meet the needs of a wide range of students.

Conclusions

Nationally, UDA is recognized as a powerful tool to support the needs of the diverse student population through integrated accommodations. The reauthorization of Individuals with Disabilities Education Act (IDEA, 2004) highlights the importance of applying the principles of universal design to tests for students with disabilities. In addition, while serving as the U. S. Secretary of Education, Roderick Paige addressed the Council of Chief State School Officers and cited the need to develop universally designed assessments to meet the demands of inclusive testing programs as specified in the *No Child Left Behind Education Act of 2001* (personal communication, June 27, 2003). Across the country, research and development efforts are emerging in response to the call for accurate and meaningful measurement of the knowledge and abilities of students with diverse needs in general assessment systems.

In this paper, I presented the theoretical basis for applying the principles of universal design to assessments. In addition, I described the process of designing an assessment device that aligned with the principles of universal design with the intention of offering insights to test developers, state assessment directors, and other researchers concerned with supporting the needs of all students in the testing process. The cornerstone of applying the principles of universal design to assessment is the elimination of inherent test characteristics that differentially influence student performance in the tested domain. The fairness of a test is jeopardized when the test takers are not provided comparable opportunities to demonstrate their abilities in the tested construct. Although external accommodations are designed to reduce these sources of test bias, the delivery and assignment of these test changes may introduce additional sources of construct-irrelevant variance that further influence the validity of the interpretations and uses of test results. Considering the principles of universal design in test development reduces the likelihood of test bias. By incorporating design features to support the needs of the diverse student population, UDA increases the participation of students with disabilities and English-language learners in general education assessments. As such, UDA offers a mechanism for meeting the legislative requirements of the *No Child Left Behind Education Act of 2001*.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Bennett, R. E. (2002). Inexorable and inevitable: The continuing story of technology and assessment. *Journal of Technology, Learning, and Assessment, 1*(1). Available from http://www.jtla.org

Burk, M. (1998). *Computerized test accommodations: A new approach for inclusion and success for students with disabilities*. Washington, D.C.: A. U. Software.

Center for Universal Design. (1997). *What is universal design?* NC State University: The Center for Universal Design.

Dolan, B. (2000). Universal design for learning: Associate editor's column. *Journal of Special Education Technology, 15*(4), 47–51.

Dolan, R. P., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *Journal of Technology, Learning, and Assessment, 3*(7). Available from http://www.jtla.org

Elliott, S. N., Kratochwill, T. R., & McKevitt, B. C. (2001). Experimental analysis of the effects of testing accommodations on the scores of students with and without disabilities. *Journal of School Psychology, 39*(1), 3–24.

Fuchs, L. S. & Fuchs. D. (1999). Fair and unfair testing accommodations. *School Administrator, 56*(10), 24–29.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C., Binkley, E., & Crouch, R. (2000). Using objective data sources to enhance teacher judgments about test accommodations. *Exceptional Children, 67*(1), 67–81.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Helwig, R., Rozek-Tedesco, M. A., & Tindal, G. (2002). An oral verses a standard administration of a large-scale mathematics test. *Journal of Special Education, 36*(1), 39–47.

Helwig, R., Rozek-Tedesco, M. A., Tindal, G., Heath, B., & Almond, P. J. (1999). Reading as an access to mathematics problem solving on multiple-choice tests for sixth-grade students. *Journal of Educational Research, 93*(2), 113–125.

Helwig, R. & Tindal, G. (2003). An experimental analysis of accommodation decisions on large-scale mathematics tests. *Council for Exceptional Children, 69*(2), 211–225.

Hickman, J. A. (2000). Individualized education plan (IEP). In C.R. Reynolds & E. Fletcher-Janzen (Eds.), *Encyclopedia of Special Education* (pp. 939–940). New York: John Wiley & Sons.

Higbee, J. L. (2001). Implications of universal instructional design for developmental education. *Research and Teaching in Developmental Education, 17*(2), 67–70.

Hopper, M. F. (2001). The implications of accommodations in testing students with disabilities. (ERIC Document Reproduction Service No. ED 463 627).

Ketterlin-Geller, L. R. (2003). *Establishing a validity argument for universally designed assessments*. Unpublished Doctoral Dissertation, University of Oregon, Eugene, OR.

Ketterlin-Geller, L. R., Alonzo, J., & Tindal, G. (2004). *Use of Focus Groups to Inform the Construction of a Universally Designed Mathematics Test* (Tech. Rep. No. 29). Eugene, Oregon: University of Oregon, College of Education, Behavioral Research and Teaching.

Lowe, P. A. & Reynolds, C. R. (2000). Individuals with disabilities education act (IDEA). In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Encyclopedia of Special Education* (pp. 940–948). New York: John Wiley & Sons.

Mace, R. L., Hardie, G. J., & Place, J. P. (1991). Accessible environments: Toward universal design. In W. E. Preiser, J. C. Vischer, & E. T. White (Eds.). *Design intervention: Toward a more humane architecture.* New York: Van Nostrand Reinhold.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5–11.

National Center on Educational Outcomes. (2003). *Special topic area: Accommodations for students with disabilities.* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 30, 2003, from the World Wide Web: http://education.umn.edu/NCEO/TopicAreas/Accommodations/Accom_topic

Pisha, B. & Coyne, P. (2001). Smart from the start: The promise of universal design for learning. *Remedial and Special Education, 22*(4), 197–203.

Rudner, L. M. (1998). *An on-line, interactive, computer adaptive testing tutorial.* Retrieved on September 15, 2003 from http://ericae.net/scripts/cat

Shriner, J. G. & Destefano, L. (2003). Participation and accommodation in state assessment: The role of individualized education programs. *Council for Exceptional Children, 69*(2), 147–161.

Stansfield, C. W. (2002). Linguistic simplification: A promising test accommodation for LEP students? *Practical Assessment, Research & Evaluation, 8*(7). Retrieved on April 30, 2003 from http://ericae.net/pare/getvn.asp?v=8&n=7

Story, M. F., Mueller, J. L., & Mace, R. L. (1998). *The universal design file: Designing for people of all ages and abilities.* NC State University: The Center for Universal Design.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 7, 2003, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.htm

Thompson, S. & Thurlow, M. (2002). *Universal design assessments: Better tests for everyone!* (Policy Directions No. 14). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved April 7, 2003, from the World Wide Web: *http://education.umn.edu/NCEO/OnlinePubs/Policy14.htm*

Thompson, S., Thurlow, M., & Moore, M. (2002). *Using computer-based tests with students with disabilities* (Policy Directions No. 15). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved January 13, 2003, from the World Wide Web: http://education.umn.edu/NCEO/OnlinePubs/Policy15.htm

Tindal, G., Anderson, L., Helwig, R., Miller, & Glasgow, A. (2000). *Accommodating students with learning disabilities on math tests using language simplification.* Eugene, OR: RCTP.

Tindal, G. & Crawford, L. (2003). *Technology applications for students with disabilities: Using learning assessments to design effective programs.* Eugene, OR: Behavioral Research and Teaching.

U.S. Department of Justice. (2003). *ADA Regulations and Technical Assistance Materials.* Retrieved April 30, 2003 from http://www.usdoj.gov

Author Biography

Leanne Ketterlin-Geller is an Assistant Professor at the University of Oregon where she also serves as the Director of Research for Behavioral Research and Teaching, a not-for-profit education research organization. Dr. Ketterlin-Geller's research focuses on the development of effective assessment procedures and valid decision-making systems for students with disabilities and culturally and linguistically diverse students in the general education curriculum. Much of her work is centered on using technology to provide a flexible assessment system for supporting students' needs.

J·T·L·A

The Journal of Technology, Learning, and Assessment

www.jtla.org

Technology and Assessment Study Collaborative
Caroline A. & Peter S. Lynch School of Education, Boston College