SECTION VI
Student Preparation for College

Your Key to Evaluation: The Computerized Adaptive Testing Version of the Transition Competence Battery for Deaf Adolescents and Young Adults

Cheryl Davis, WROCC Coordinator
Western Oregon University
Monmouth, OR

Annette Leonard, Teaching Research Division
Western Oregon University
Monmouth, OR

Abstract

Assessment data should serve as the cornerstone of the transition and rehabilitation process, guiding the provision of service offerings. The instruments that are employed and the data derived should relate specifically to the content of the transition process. This type of information will have relevance for instruction (i.e., determining what the individual needs to be taught) and for planning the transition from the school to the community. Unfortunately, psychometrically sound assessment instruments developed specifically for deaf adolescents and young adults are rare. This paper describes the development and current state of research on one such instrument: The Transition Competence Battery.

For too many years teachers, parents, and professionals working with deaf adolescents and young adults have faced the difficulties of adapting materials developed for students without hearing losses for their deaf students. This creates a 'Catch 22' for educators: 1) instruments not normed for deaf students become invalid for the population once administration procedures are altered, thus providing inaccurate information about the deaf student's status (but they must be altered in order to administer them); 2) hearing is assumed to be present in these materials, and so testing of skills particular to deaf student's life experiences is necessarily absent; 3) the language used in materials developed for hearing students is often inaccessible to deaf students, resulting in test scores that reflect a deaf student's English comprehension and *not* their mastery of the subject. Obviously, this serves neither the students nor the educators.

The Transition Competence Battery (TCB) is a unique assessment tool designed to measure the transition skills of deaf high school adolescents and young adults who plan to enter the work force, job training programs, or two-year colleges. This segment of the deaf, high school population is both the largest and in need of the most services to succeed in their transition to adult life. Development and research of the TCB began 16 years ago. When the project was initiated in 1986, a review of the pertinent literature and existing curriculum demonstrated some significant areas of weakness pointing to the fact that deaf students' transition needs were being overlooked and underserved. The factors that provided the impetus for the TCB's creation are that tests are not often written or standardized for deaf students, linguistic access in assessment measures has not been a design consideration, and that other transition assessment instruments are not inclusive of deaf students' life experiences.

Funded by a grant from the Federal Office of Special Education Programs, the TCB was stringently developed with the goal of producing a nationally standardized assessment tool. The defining characteristic of the TCB is its accessibility to students who rely on sign communication.

Transition Competence Battery

The first phase began with a nominal group technique (NGT) of 18 professionals and consumers. Seeking input from teachers, administrators, VR counselors, parents, employers, and deaf adults, two questions were asked: 1) What are the five most important independent living skills Deaf students should have? 2) What are the five most important employment skills Deaf students should have? Project staff then grouped the responses into three areas each within two broad domains. The Employment Domain includes job-seeking skills, job related social skills, and work adjustment skills. The Independent Living Domain includes health and home skills, money management skills, and community awareness skills.

A national survey was then conducted asking respondents to rate the skills developed in the NGT according to 1) the skill's importance in the successful transition of deaf students and 2) the presence of the skill, i.e., do most students perform the skill well or is it an area of weakness for most students? From these ratings, the skills were prioritized and those rated as most important and least often present were selected for inclusion in the test battery. A second group of professionals were convened to begin item writing. Translating these skill areas into questions and then into a test required careful consideration of issues such as, target population's literacy; what language should be used; how the questions should be formatted; whether or not the use of graphics would detract from the test's utility; and whether or not the test should be

time limited. All items are presented on video in conceptually accurate Pidgin Sign English (PSE) in a 3-option multiple-choice format. More than 900 items were developed to be considered for inclusion in the final product. Two hundred items were selected from this initial item pool as appropriate for the 6 subtests.

Nationwide standardization of the TCB consisted of traveling to schools for the deaf and mainstream programs across the US with the 6 videotape subtests. The test sessions were conducted in small groups (generally 6-10 students) watching the test signed in conceptually accurate Pidgin Sign English on a large TV monitor. Standardized instructions were presented in PSE on the videotape of the first subtest. Items were presented in a multiple choice, three option format with a maximum amount of time allotted for responding to each item (the test was not meant to be timed, however, in order for the administration to be standardized, there needed to be a fixed time limit on response time). After the question and response options were presented to the students on the video, they had the opportunity to read a written version of the question in a test booklet. They responded on a separate piece of paper by circling the correct response (A, B, or C). They were also provided scratch paper to use for math items. Calculators were not allowed. Each test required approximately 30 – 45 minutes to administer. Because of the length, some sites broke the testing into two 3-hour days.

Fourteen programs and 230 students participated in the project. In order to be retained in the battery, items had to possess a point-biserial correlation of at least .2, reach a moderate level of difficulty (i.e., 40 – 80 percent of students answering item correctly), and the signed version of the item had to be judged conceptually accurate by project staff. From the 200 items field tested, eighteen items were deleted from the final version of the battery.

The proportion of items answered correctly by students ranged from a low of .51 on the Money Management subtest to a high of .73 on the Job-related Social Skills subtest. The tests' internal consistency reliability indices ranged from .67 on the Money Management subtest to .86 on the Job Seeking Skills subtest. In general, an index of .75 to .80 is regarded as acceptable for group administered screening instruments (Bolton, 1987; Salvia & Ysseldyke, 1986). All subtests exceeded this standard except the Money Management. This subtest was the shortest (only 20 items) and also included several functional math items, and therefore may have involved more guessing on the students' part. Test-retest correlations were strong (.77 to .90), again with the exception of the Money Management subtest (.61). This supports our hypothesis that guessing was a significant factor in students' responses on this subtest. Complete information on the reliability and validity analyses of this project can be found in Bullis and Reiman (1992).

Mini Transition Competence Battery

Even though the results of the development of the TCB were very positive, several improvements to the test could be made. With group administrations of the videotape assessment, some students must wait for others to finish answering each question before the next video question is presented. Thus, the slower students may experience peer pressure to complete items quickly; faster students may become bored and be distracting to the other students. Even though item statistics surpassed acceptable levels, our later research comparing group with individual administration showed improved scores and psychometrics in the individual administrations. Test administrators gained a greater intuitive feel of the individual's skills and capabilities and would prefer this approach if it were not for 6 to 8 hour per pupil administration time. Finally, the amount of time required for testing was simply prohibitive for many sites.

In response to these concerns, a second line of research began to develop the Mini TCB. Using the items developed for the TCB and the data collected in the standardization research, Item Response Theory (IRT) parameter estimates were computed using three logistic models. All six subtests were essentially unidimensional for adequate IRT modeling. Items were screened utilizing a three- parameter model (item difficulty, item discrimination, and guessing). Using this model, we identified 48 items that best measured performance at the proficiency (screening) level of ability.

Five hundred and ninety three students participated in the field test of the Mini TCB, including 133 four-year college students making up a comparison group. Concurrent validity study subjects took the Mini TCB and the full test battery 2-4 weeks later. Correlations between the Mini TCB and each of the subtests ranged from .73 to .86. Reliability study subjects took the Mini TCB once, and then again 2-4 weeks later. The test-retest index was .75. Students in the construct validity study took both the Mini TCB and the 6 subtests of the TCB. In the planned comparisons, we found that the college students (students definitely *not* included in the target population) performed significantly higher on the Mini and the six subtests than mainstream, residential, and community college students. Complete information on the reliability and validity analyses of this project can be found in Bullis, Reiman, Davis, and Reid (1997) and Bullis, Reiman, Reid, and Davis (1995).

Taken together, these results provide strong support for the psychometric characteristics of the Mini TCB, suggesting that it would be a suitable and important instrument to use with adolescents who are deaf and in transition. This screening tool requires approximately 45 minutes to administer. Large groups of students can be screened with this tool, and the more intensive test battery need be administered only to those students not passing the screening.

Computerized Adaptive Testing version of the TCB (CAT-TCB)

The Mini TCB has been well received by users in the field. The reduction of time required for testing has made it much more accessible to school programs. Since the beginning of the research on the TCB, though, there has been a huge increase in computer use, and many have requested a computerized version of the test rather than videotape. A computerized version of the test (e.g., CD ROM) would be useful in numerous ways: a computer program installed using a CD ROM is less susceptible to the wear and tear to which a videotape is exposed; because the test would be administered individually and all students receive a different combination of items, peer pressure would no longer an issue; and because students would be able to respond directly on the computer, reports could be compiled and hand scoring would no longer be necessary.

Until recently, the technology simply did not exist that allowed for the high quality video required in the instruction of deaf students to be shown on computers. Both the quality of the video and the size of the video files made a computerized version of the test impossible. Fifteen years later, 80 Gigabyte hard drives are available on laptops, processor speeds and increased video memory allow for smooth presentation of video, and storage options such as CD and DVD make portability possible as well. In addition, adaptive testing (which tailors the test to the individual student) is an exciting possibility in a computerized version of the test.

Item Response Theory (IRT) is a statistical technique that allows us to determine, among other parameters, the difficulty of each item based on the information gained from our nationwide standardization data. The information gained from the IRT analysis has been used to write a computer program that, based on the student's correct or incorrect response to the current question, will choose the next appropriate item to present to that student. An incorrect response leads to a less difficult item; a correct response leads to a more difficult item. This process, known as Computerized Adaptive Testing (CAT) results in 1) increased efficiency in testing since fewer item responses are required to maximize confidence in the estimated ability level, 2) greatly reduces total testing time–*we can assess students in about 45 minutes across the entire test battery*, 3) individual pacing so that students can work at their own speed, 4) an increased interest level for the student as only items of appropriate difficulty are used, and 5) immediate scoring and reporting via a computerized printout of the results. As with the TCB, the results of the testing can be used in developing Individualized Education Plans and Individualized Transition Plans.

Rather than six different subtests, the CAT-TCB is presented in two domains: Employment and Independent Living. Students view standardized instructions on the computer which include information about the research project, how to use the navigation buttons in the program, how to respond to the items, and examples.

Students see the item stem and response options signed in PSE first. Next, the response screen appears. It includes a written version of the question and the three response options. Students can choose to view that question over, or pick a response and choose to continue. Responding is a two-step process that allows the student to change his or her mind and select a different answer before continuing.

The reporting function allows for both individual and group data to be compiled for the instructor. Students and instructors have responded very positively to this new version of the test. Testing in both domains takes about 45 minutes to an hour.

Summary

Field testing is nearing completion and analysis will take place in the Fall of 2002. The CAT-TCB is being analyzed by comparing student performance on it with the Mini TCB, and via formative evaluation information collected from participants. Updates will be available on the CAT-TCB website at <http://www.wou.edu/nwoc/cattcb/cat-tcbhome.htm>. A commercial publisher will be sought. The TCB and the Mini TCB are currently available from James Stanfield Publishing Company at <http://www.stanfield.com>.

References

Bolton, B. (1987). *Handbook of measurement and evaluation in rehabilitation*. Baltimore: Paul H. Brookes.

Bullis, M., & Reiman, J. (1992). Development and preliminary psychometric characteristics of the Transition Competence Battery for Deaf Adolescents and Young Adults. *Exceptional Children, 59,* 12-26.

Bullis, M., Reiman, J., Davis, C., & Reid, C. (1997). National field-test of the mini version of the Transition Competence Battery for Deaf Adolescents and Young Adults. *The Journal of Special Education, 31,* 347-361.

Bullis, M., Reiman, J., Reid, C., & Davis, C. (1995). Development and preliminary psychometric characteristics of the "mini" version of Transition Competence Battery for Deaf Adolescents and Young Adults. *Assessment in Rehabilitation and Exceptionality, 2,* 179-196.

Salvia, J., & Ysseldyke, J. (1988). *Assessment in special and remedial education*. (4th ed.). Boston: Houghton Mifflin.