TBAC

A Brief Manual of Instructions for the use of the

TEST OF BASIC AUDITORY CAPABILITIES, MODIFICATION 4

(Revised, August, 2009)

1. Introduction.

a. ***History of the TBAC.*** The original Test of Basic Auditory Capabilities (TBAC) was developed in the early 1980's as an off-shoot of a program of studies of auditory pattern discrimination. Those studies were conducted by Chuck Watson and colleagues at the Central Institute for the Deaf in St. Louis and at the Boys Town Institute for Communication Disorders in Children in Omaha (now the Boys Town National Research Hospital).

The primary reason for developing this test battery was that individual differences in the ability to detect or discriminate changes in complex auditory stimuli, such as sequential tonal patterns, or in "profiles" (complexes consisting of multiple simultaneous tones studied intensively by David Green (1987) and colleagues) had been found to be unexpectedly large. Thresholds in dB, Hz or milliseconds for the discrimination of complex sounds were often found to be not only many times larger, but far more variable among listeners, than were those for the detection or discrimination of simpler sounds, as, for example, the individual tone pulses so commonly used in earlier auditory research. Smaller individual differences among so-called "normal hearing listeners" had previously been written off as largely matters of measurement error, or the unreliability of the test procedures. With the more complex stimuli, these differences were not only larger, but were found to be reliably measureable characteristics of individual listeners.

Many investigators, studying a variety of complex sounds, had reported that the range of individual differences among their listeners was often as large as the effects they were attempting to study. After repeated reports of that sort, we decided to develop a test battery for the explicit study of the ways that individuals differ in their auditory abilities. While there had been earlier studies of individual differences in auditory abilities, none had been conducted since the advent of Signal Detection Theory (Green and Swets, 1966). The older studies were designed without recognizing the need to control the effects of response bias, or to provide discrete measures of bias and sensitivity. In other words, previous test batteries had generally employed "criterion confounded methods." Whether this in fact made a great difference in the measurements obtained may reasonably be questioned, nevertheless the potential value of criterion-controlled methods was one of the motivations that led to the development of the TBAC. Following some preliminary work with a 28-subtest battery (Johnson, Watson and Jensen, 1987) the first version of the TBAC was recorded, a battery with eight subtests, six using single tones or tonal patterns and two with speech sounds. The Johnson et al. study had demonstrated that individual listeners often have specific areas of excellent or poor auditory acuity (as on all the tests involving frequency discrimination) but their work also showed that much larger numbers of listeners (than the 24 tested in that study) needed to be included, if auditory abilities were to be identified by means of contemporary multivariate statistical techniques (e.g. factor analyses, structural equation modeling).

b. ***Contemporary research on individual differences in auditory abilities.*** While the TBAC was originally developed to simply document the range of auditory abilities on a somewhat arbitrarily selected set of auditory tasks, recent test-battery research has taken a more principled approach. It has been motivated by three basic questions: 1) Are there discrete auditory abilities? 2) If so, what are they? And 3) If there are discrete auditory abilities, how are they distributed in populations of normal-hearing and hearing-impaired adults. An additional question that has been addressed by test-battery research has concerned the inter-relation among the various auditory abilities, particularly that between measures of temporal and spectral acuity, and the ability to understand speech. These topics are discussed at length in articles listed in the reference section of this manual. In brief, it has been shown that there may be at least four discrete auditory abilities (see Figure 1 below, from Kidd et al., 2007), plus a general auditory ability ("Auditory G"), all of which were found to vary independently of general intelligence (IQ), in a population of healthy young adults. It has also been found that measures of spectral and temporal acuity obtained with laboratory test sounds are very poor predictors of speech recognition under difficult listening conditions. Instead, listeners appear to have varying levels of the ability to recognize familiar sounds (speech or environmental sounds) on the basis of partial information (as, when masking noises obscure some portions of the stimuli).

c. ***The tone tests.*** Since it is well established that listeners make use of subtle spectral and temporal properties to identify sounds and sound sources, both speech and nonspeech, the tests should include measures of temporal and spectral acuity, using both simple and complex sounds. Therefore tests 1-3 measure the abilities to discriminate single tones on the basis of frequency, duration and intensity. Tests 4-6 measure what may be considered to be higher-level abilities, in the sense that they place greater demands on selective attention and memory than do single-tone tests. The fourth test measures the ability to detect changes in the rhythm of a four-pulse series, the fifth the ability to detect the presence of an extra tone in the temporal middle of a one-half-second (word length) ten-tone pattern, while the sixth test requires that the listener discriminate between four-tone sequences on the basis of the order of the middle two tones, either high-low, or low-high. The frequency for these tests, or the mean frequency, is 1000 Hz, chosen because it is toward the middle of the spectrum of audible frequencies, it is less likely than much higher or lower frequencies to be distorted by reproduction technology, and it is low enough to be audible to most listeners with mild–to-moderate sensorineural hearing loss.

d. The speech tests. The first of the two speech tests (test No. 7) is modeled on the final tone test (test No. 6), in that four syllables are presented, /fa/ta/ka/pa/, and the listener's task is to discriminate between the syllable sequences on the basis of the order of the middle two syllables, either /ta/ka/ or /ka/ta/. This task is made more difficult by reducing the duration of the vowel (/a/) in the sequences. The second speech test (subtest No. 8) requires that the listener select from among three alternative syllables the one that was presented in a noisy background. Note that the TBAC, Modification 4 (hereafter, the "TBAC-4") employs a considerably more reliable syllable test than did the original TBAC, as discussed in a later section.

e. ***Selection of Subtests for the TBAC-4.*** The tests just described were included in the TBAC-4 battery because they each load strongly on one among the four auditory abilities described by Kidd et al. (2007), and also because they jointly provide a reliable estimate of the General Auditory Ability (G_A). The discrete ability that is not well estimated by any of the eight TBAC-4 subtests is that associated with the detection of the depth of modulation of samples of sinusoidally amplitude modulated (SAM) noises. SAM tests were not included in this battery, as it was already slightly over an hour in length, and because several of the tests that are included clearly place strong demands on the listeners' temporal acuity. SAM tests may be added to later versions of the TBAC if there is sufficient interest in including those measures.

f. ***Reliability of the TBAC***. The TBAC has excellent psychometric reliability, given that it has relatively few trials on each of the individual subtests. Christopherson and Humes (1992) examined the reliability of the original eight TBAC subtests. The original TBAC was administered multiple times to the same listeners and was found to be reliable; Cronbach's alpha values were above 0.7 for all but the Syllable Identification subtest, which had a value of 0.58 (note that for this reason the original syllable identification task has been replaced with a much more reliable syllable subtest in the present battery). Performance on all subtests changed little over six repeated administrations of the TBAC. The reliability of all subtests was estimated a second time by Kidd, Watson and Gygi (2007) as part of a study of individual differences in auditory abilities, in which 338 listeners were tested on the original TBAC tests, plus eleven additional subtests that were added to the battery. Kidd et al., estimated reliability by means of a split-half procedure, rather than the repeated-test metjhod used by Christopherson and Humes. Utilizing a resampling strategy (Good, 2006), correlations for 1000 randomly selected pairs of split halves were computed for each subtest, and reliability was computed using the Spearman-Brown prediction formula applied to the mean of that sample of 1000 correlations. The results are shown in Table 1.

Consistent with the findings of Christopherson and Humes, reliability coefficients for all of the first eight subtests were above 0.7, again with the exception of the Syllable Identification subtest. Reliability for the newer subtests were in the same range, with only three of the subtests falling below 0.7 (with the lowest coefficient at 0.501 for the original TBAC Syllable Identification test). Because of the relatively low reliability of the original Syllable Identification test, it was replaced in the TBAC-4, by the “Non-word test” from the TBAC-E battery (here re-named the Syllable Recognition Test), for which the coefficient of reliability was 0.787. Thus, the reliability coefficients of all TBAC-4 subtests are quite high, given their relatively brief durations. As with many behavioral measures, performance on these tests may be influenced by factors other than the abilities they were designed to estimate (such as extreme deficits in general intelligence, low motivation, fatigue, etc.).

Table 1. Split-half reliability coefficients for the 19 subtests in the expanded TBAC, including the eight tests from the original TBAC and eleven tests added by Kidd, Watson and Gygi (2007).

Subtests in original TBAC	*Reliability*
Pitch*	0.819
Loudness	0.878
Duration	0.755
Pulse train	0.816
Embedded tone	0.723
Temporal Order (tones)	0.807
Syllable Sequence	0.766
Syllable Identification	0.501

Subtests added in TBAC-E (expanded)	*Reliability*
Rippled Noise	0.759
SAM Noise, 20 Hz	0.815
SAM Noise, 60 Hz	0.702
SAM Noise, 200 Hz	0.717
Gap Discrimination	0.560
Gap Detection	0.612
Nonword Recognition (CVC syllables)	0.787
Word Recognition	0.635
Sentence Recognition	0.795
Environmental Sound Recognition	0.827

*Subtests in bold font are those included in the TBAC, Mod.4.

The reliability of repeated administrations of the tests and the stability of mean performance, as shown by Christopherson and Humes, suggests that greater reliability than that reported in Table 1 could be achieved, if desired, by simply giving an individual test 2-3 times and computing the mean score. It should be noted, however, that this version (TBAC-4) does differ from the earlier tests in that trial-by-trial feedback is presented to the listeners. It is thus possible that on repeated administrations the listeners might begin to learn the sequences of correct responses, on the basis of that feedback. Versions of the test without trial-by-trial feed back can be provided.

g. ***Recent Research on the TBAC.*** The bar graph in Figure 1 shows the results of a factor analysis applied to the 19 subtests of the expanded TBAC-E battery (the tests shown in Table 1). This figure shows the factor loadings of each subtest on each of four factors, or underlying auditory abilities. These four abilities, together with the General Auditory Ability ("Auditory G" or G_A) discussed in the following paragraph, provide a reasonable account of performance on the nineteen subtests of the expanded TBAC battery used by Kidd et al. (2007). In this figure the tests are grouped on the basis of the single factor on which each test had the highest loading, and are ordered within each factor by the strength of those loadings. Note that most tests have from small to fairly strong loadings on more than one factor. In other words, performance on each test appears to be influenced by more than one underlying ability, although a few tests come fairly close to being "pure" measures of a single factor. Among the original eight TBAC subtests, Intensity and Duration fall in Factor 1 (consistent with that factor representing sensitivity to total stimulus energy). The original Syllable ID test also falls in that factor, but it has a moderately strong secondary loading on the Familiar Sounds Factor, which includes most of the other speech recognition tasks and also a test of the ability to identify common environmental sounds. This, along with its relatively low reliability, suggested that the original Syllable ID task (test items like, "you will mark *ooz*, please," presented in cafeteria noise) was not a particularly good general test of speech recognition. The several other speech tests in the TBAC-E are both more reliable and more highly correlated with each other. One of those tests, that requires CVC identification in a Gaussian noise background (labeled "NonWords" in Figure 1) has replaced the older TBAC Syllable ID test in the TBAC-4. The Pitch-Discrimination, Temporal-Order-for-Tones, Embedded-Tone and Syllable- Sequence tasks all fall in the factor labeled "Pitch and Time." The common element among these tasks is that they all require listeners to detect spectral or temporal changes in relatively unfamiliar laboratory-generated sounds. (It might seem that the syllable sequence task is a familiar one, but the way that task is made difficult is by deleting temporal segments of the /a/ vowel, in the /fa/ta/ka/pa/ sequences, in order to make the order of the syllables more difficult to recognize. This manipulation yields rapid sequences that have a very unnatural quality, as can be verified by listening to them.) As mentioned earlier, the Amplitude Modulation factor is not included in the TBAC-4.

Kidd et al. (2007) examined several structural-equation models as candidate explanations of the performance of 338 college-age listeners with normal audiograms on the 19 subtests in the TBAC-E battery. The best fitting of these was a model that assumes performance on each of the tests to be accounted for by a specific combination of the four discrete auditory abilities (Fig. 1), plus some level of influence of a general auditory ability, or "Auditory G". Auditory G accounts for about as much variance as do the four individual auditory abilities. It should also be emphasized that these listeners' general intellectual abilities were estimated from their Scholastic Aptitude Test (SAT) scores (Math and Verbal) and that *those measures accounted for little or no variation in* *the TBAC scores*. Since it has recently been shown that SAT scores provide reasonably reliable estimates of full-scale IQ (Frey and Detterman, 2004; Beaujean et al., 2006) this supports the hypothesis that both general and specific auditory abilities have little or no association with (are statistically independent of) general intelligence.

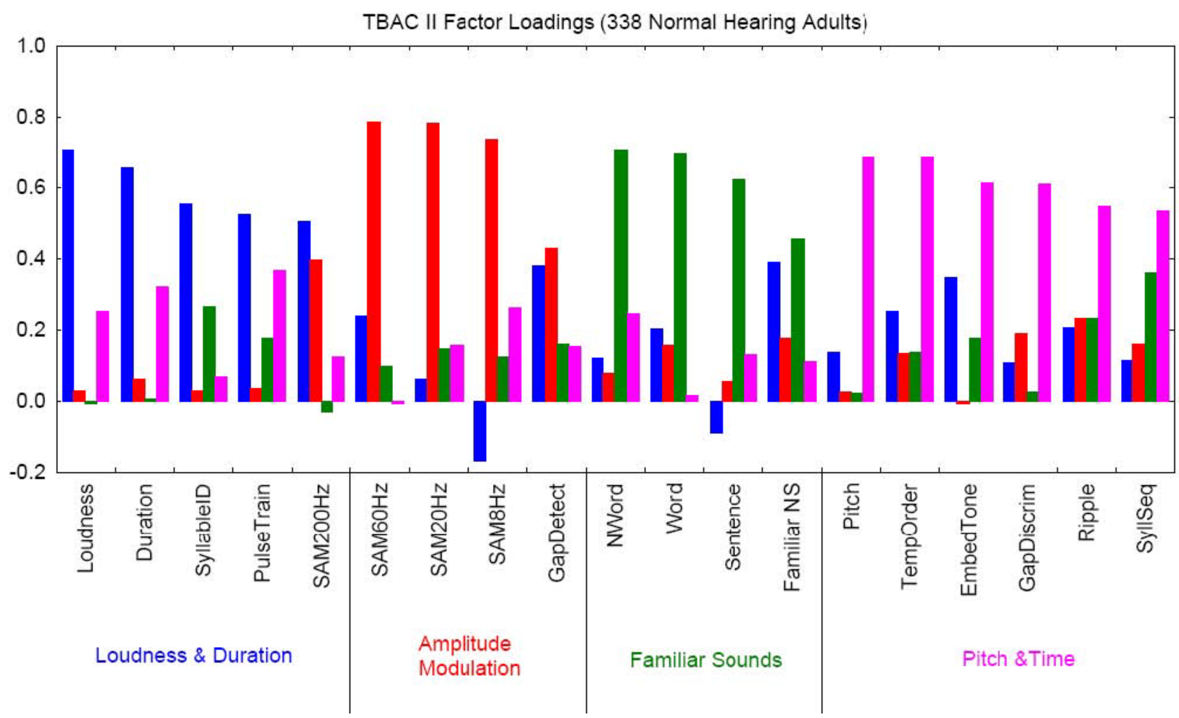

Figure 1. Four-factor analysis of the performance of 338 college students with normal audiograms, on the 19 subtests of the expanded TBAC-E battery (Kidd, Watson and Gygi, 2007).

h. ***Purchasing the TBAC and individual Scoring Units (SUs).*** Go to the CDT purchasing page at **http://www.comdistec.com/new/catalog.html**. On that page you will be able to purchase the TBAC software and additional Scoring Units, using either a major credit card or PayPal. After you have made your purchase, you will be sent instructions for downloading the TBAC.

Eight Scoring Units are included with the TBAC software. Additional units are available at $4.00 each. When the TBAC is administered to one or more subjects (depending on the number of scoring units purchased) the results are uploaded (anonymously) to CDT and the scores for each subject are placed on the users C Drive, in a directory called “TBAC_SCORES.” The scores provided in that directory include percent correct on each subtest, the percentile ranking relative to a sample of 340 college students with normal pure-tone audiograms, and an estimate of the listener’s threshold in milliseconds, Hz, or dB, as appropriate for each subtest.

2. Using the TBAC, Mod.4.

a. Required and optional equipment. The TBAC is designed to be used with a personal computer running Windows XP. The program occupies 498 megabytes (which is mostly .wav files). A pair of good-quality headphones is also required. Although good results can be obtained without expensive laboratory-grade equipment, lower-quality equipment may introduce sufficient noise or distortion to substantially reduce the reliability of the tests.

Headphones. Circumaural headphones (ones that surround the ear) are recommended to help attenuate distracting sounds. However good-quality supra-aural headphones (which rest on the ear, but do not surround it), such as the Grado SR60 and SR80, can also work well in a quiet environment. Inexpensive earbud headsets should be avoided, but high-quality earbuds that fit snugly in the ear canal can produce good results. Many manufacturers (e.g., Grado, Sony, AKG, Sennheisser) make very good quality headphones for less than $100.

Sound card. Most popular sound cards from major manufacturers will produce acceptable results. Many computers, however, deliver sound through audio chipsets included on the computer's motherboard. These chipsets can also produce acceptable results, but the likelihood of significant noise or distortion is generally greater with this type of sound system. Many problems can be detected by listening carefully to the recorded sounds (such as the introduction and spoken instructions). If there are audible noises (such as a "hum" or a "hiss"), or if any tones do not sound like single pure tones (e.g. have a buzzing quality) the TBAC results may be compromised. Note: Be sure to turn off any special effects (e.g., reverberation, simulated 3D sound) and set any tone controls or equalization to a neutral position to minimize disparities in the intensities of different frequencies.

b. ***Opening the Application***. When the TBAC application is first opened, the name of the administrator of the test battery should be entered in the place provided. After that is done, a window labeled TBAC Settings is displayed. This window provides an opportunity for the administrator to calibrate the level at which the test sounds will be presented and also several options for the conduct of the test, as described in the following sections.

c. ***Calibration.*** It is possible to use the TBAC with reasonable confidence without calibrating the testing system with a sound level meter and the specific earphones to be utilized, however a band-pass noise is provided for this purpose. The calibration noise is about one minute long and it is initiated from the administrator's console, labeled "TBAC Settings" (right click on the User ID field, prior to starting a test). To turn on the calibration noise, select "*Begin Calibration*." If a sound-level meter is used for this purpose, the level of the calibration noise at the earphone's output should be set to 75 dB, SPL. If calibration equipment is not available, valid and reliable results can still be achieved by instructing a listener with normal auditory sensitivity (no hearing loss) to listen to the calibration noise over earphones and to set the level to, “That which you find comfortable, with the understanding that you will have to hear some very subtle differences between test sounds presented at this level.” This instruction is generally successful in convincing listeners to adjust the calibration noise so that sound levels for the test will be in the 60-80 dB range. Throughout that range discrimination performance on tests similar to those in the TBAC tend to be essentially constant for listeners with normal sensitivity. ***It is vital that the administrator also listen to the calibration noise at the level selected, if this latter method of selecting the level is utilized. This is necessary in order to avoid the possibility of exposing listeners to excessive sound levels.***

d. ***Administrator's options.*** As mentioned in the preceding sections, the window called “TBAC Settings” should be accessed by the administrator prior to starting the TBAC. In this window there are several options available. These are:

1) Choice of whether the TBAC subtests will be automatically presented in their standard order (the default mode), or instead be selected from a menu prior to each test. Click on the option labeled either “System” (no menu), OR “User” (menu display).

2) Trial “Pacing” options. The test trials are either presented automatically three seconds after the listener's response to the previous trial (“System Paced” which is the default mode) , or the trials can be “User Paced”, requiring the user to initiate each tiral after responding to the previous trial ; and 3) If the “User Paced” option is selected, either the space bar or a key click can be chosen as the way to initiate each successive trial.

e. ***Instructions to the listener.*** The TBAC has, from its earliest versions, been almost a “hands-off” test battery, in the sense that recorded instructions are presented to the listener at each step. In the TBAC-4, these instructions have been slightly revised and are augmented with screen messages, but they are very similar to those of the original test. The most distinctive feature of these instructions is that they are very repetitive and listeners often comment that they may not need to be told repeatedly that they must select the “Test sound that is different from the standard, either T1 or T2, on each trial” and that they should select their best guess when they are uncertain. The goal of the TBAC was to create a series of auditory tests on which subjects would never, or hardly ever, become confused about what they are supposed to do. Experience has shown this to have been achieved; only a very few listeners, out of thousands tested, have ever failed to respond to every trial as requested. *Thus the administrator is urged to resist the impulse to skip the instructions and allow the listener to simply begin with the first subtest.* While that can be done, doing so may introduce variance that invalidates the percentile scores and threshold estimates provided at the completion of the test battery. These scores are based on a group of 338 young adult listeners who listened to all of the instructions and completed all of the practice trials. Moreover, the good reliability of the tests is likely due, in part, to its "ritualized" format. The listener is introduced to the test structure with the frequency discrimination subtest. Since large changes in frequency are extremely salient to listeners the test procedures are readily learned with that test. All the other tests (except for Subtest 8) use that identical format, and thus the listeners are quite expert at it by the time they complete the first three tests with single tones and are challenged by the more complex sounds in the later subtests. If the administrator feels the need to say something to the listener about the test before it begins, it is usually sufficient to assure them that, although it may seem a bit boring at times, if they try to attend carefully to the sounds that their scores will give them an accurate estimate of their auditory abilities as compared to those of typical young adults with normal auditory sensitivity.

f. ***Starting the test.*** After the level has been set, either by use of a sound level meter or by setting it to a comfortable listening level, and the desired options have been selected from the "TBAC Settings" console, click on "Continue". The administrator will then be prompted to enter an identification number for the listener (use of real names is discouraged) and other demographic information including sex, age, known hearing impairment, native language and years of musical training (provision is made for additional optional fields if these are needed for a research application). When that has been done, the test may be started by again clicking Continue, and from that point on it does not require any further intervention, unless the listener (or administrator) decides that he or she does not want to complete the entire battery without a break. This is common, since the total battery requires about an hour and 10-15 minutes, depending on the pacing option selected and the listener's response latencies. The way to pause the test depends on the options selected for its administration. If the "User" Menu option has been chosen, each successive subtest is selected from the subtest list by the user, and thus a pause can be achieved by simply not selecting the next test until the listener is ready to resume. If the "Sytem" (automatic) test sequence option has been selected, it is also easy to pause by simply not responding when the practice screen for the next subtest is presented. When ready to resume, the listener merely has to click on Continue and the instructions will begin for the next subtest.

3. Performance Measures and Their Interpretation. After the test battery has been completed the administrator may request that it, and any other listeners' data files not yet scored, be submitted for scoring. Scoring is accomplished by uploading the data files to the scoring server, which then downloads the scores and other pertinent information. Three scores are presented for each test and the listener may need some assistance in interpreting them. In addition to these scores, a measure of overall auditory ability “Auditory G” is also derived from the performance on the eight subtests.

a. **P(C).** The first score is simply the percent correct (P(C)) for all of the trials on a given subtest. It may be useful to remind the listener that since seven of the subtests use a two-alternative forced choice trial structure, that chance performance would be expected to be 50% correct, while the eighth subtest uses three alternatives, thus chance is 33% for that subtest. It might be emphasized to the listener that these overall values of percent correct are virtually meaningless except in relation to the difficulty of a specific test.

b. **Percentile Rank.** The listener’s percentile rank is obtained from a look-up table, in which interpolated values have been entered for P(C) from 50 to 99%. These values were obtained from the group of 338 young adult listeners with normal audiograms, tested by Kidd et al. (2007). The performance of all other listeners (e.g. older persons with hearing impairment) can only be interpreted as they compare to that young, non-impaired, standardization group.

c. **Threshold in dB, ms, or Hz.** An estimate of the magnitude of the stimulus change required for a listener to detect that change approximately 70% of the time is also provided. This value is reported as a “threshold” (or just detectable change) in the units of the stimulus being judged. These estimates are based on computations for each decile in the standardization group of 338 listeners (see Kidd et al. (2007) for details). The estimates reported at the end of testing are obtained from a look-up table containing interpolated threshold values associated with P(C) scores from 50% to 99%.

d. **Derived measure of overall auditory ability (Auditory G).** Several alternative models were fitted to the data collected by Kidd et al. (2007). The best fitting one included four specific auditory abilities, as illustrated in Figure 1, and one general auditory ability, or “Auditory G.” It was found that the eight subtests used in the TBAC-4 battery are sufficient to estimate Auditory G, and that index is included in the scoring. The four specific abilities cannot be reasonably estimated without including more of the original 19 subtests.

References

Christopherson, L. A., and Humes, L. E. 1992. “Some psychometric properties of the Test of Basic Auditory Capabilities TBAC,” *J. Speech Hear. Res. 35*, 929–935.

Beaujean, A. A., Firmin, M. W., Knoop, A. J., Michonski, J. D., Berry, T. P., and Lowrie, R. E. (2006). “Validation of the Frey and Detterman 2004 IQ prediction equations using the Reynolds Intellectual Assessment Scales,” *Personality & Individual Differences 41*, 353–357.

Frey, M. C., and Detterman, D. K. (2004). “Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability,” *Psychol*. *Sci*. *15*, 373–378.

Good, P. I. (2006) Resampling *Methods: A Practical Guide to Data Analysis, 3rd ed.* Birkhauser, Boston.

Green, D. M. (1987) *Profile* Analysis*: Auditory Intensity Discrimination,* Oxford University Press.

Green, D. M. and Swets, J. A. (1966) *Signal detection theory and psychophysics*. New York: Wiley.

Johnson, D. M., Watson, C. S. and Jensen, J.K. (1987) Individual differences in auditory capabilities. I. *J. Acoust. Soc. Am., 81*, 427-38.

Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *J. Acoust. Soc. Am*. *122*, 418-435.

Surprenant, A. M. and Watson, C.S. (2001) Individual differences in the processing of speech and nonspeech sounds by normal-hearing listeners. *J. Acoust. Soc. Am., 110*, 2085-95.