1-1-2006

Local norms for deaf and hard-of-hearing students

Kecia A. Boncek

Local Norms for Deaf and Hard-of-Hearing Students

Graduate Thesis

Submitted to the Faculty

Of the School Psychology Program

College of Liberal Arts
ROCHESTER INSTITUTE OF TECHNOLOGY

By

Kecia A. Boncek

In Partial Fulfillment of the Requirements
for the Degree of
Master of Science and
Advanced Graduate Certificate

Rochester, New York April 5, 2006

Approved: Jennifer Lukomski
(committee chair)
Scott P. Merydith
(committee member)

R I T
School Psychology Program
Permission to Reproduce Thesis

PERMISSION GRANTED
Title of thesis Local Norms for Deaf and Hard-of-Hearing Students

I Kecia Boncek hereby **grant permissio**n to the Wallace Memorial Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: 4/13/06 **Signature of Author:** Kecia A. Boncek

PERMISSION FROM AUTHOR REQUIRED
Title of thesis _____

I _____ **prefer to be** contacted each time a request for reproduction is made. I can be reached at the following address:

_____
_____
_____

PHONE: _____

Date: _____ **Signature of Author:** _____

PERMISSION DENIED
TITLE OF THESIS _____

I _____ hereby **deny** permission to theWallace Memorial Library of the Rochester Institute of Technology to reproduce my thesis in whole or in part.

Date: _____ **Signature of Author:** _____

Running head: LOCAL NORMS FOR DEAF AND HOH STUDENTS

Local Norms for Deaf and Hard-of-Hearing Students

Kecia A. Boncek

Rochester Institute of Technology

Abstract

Deafness is a low incident disability and therefore deaf children are frequently not included in the norm group of norm-referenced tests (Braden, 1994). The necessity of *local norms* is an ongoing debate among professionals who assess deaf children (Braden, 2005). This study evaluated the utility of *local norms* for three different intelligence tests. One sample t-tests were run to determine significance differences between means for a deaf and hard-of-hearing sample compared to the norm group, for the Differential Ability Scales (DAS), Universal Nonverbal Intelligence Test (UNIT), and the Wechsler Intelligence Scale for Children- Third Edition (WISC-III). Verbal cluster and Word Definition scores on the DAS were significantly different from the published means, as well as the Vocabulary scaled score on the WISC-III. Local norms were created for these variables.

According to Sattler (2001) the purpose of a psychoeducational assessment is to obtain information about an individual to aid in educational decision-making. Many examiners use norm-referenced tests to obtain such information. Commercial norm-referenced tests provide a score that ranks the individual's performance compared to a normative sample (Ysseldyke, 2004). Since many clinicians value norm-referenced tests and use them for decision-making, individuals included in the norm sample need to be representative of the general population who will be tested using norm-referenced tests.

Although there are criticisms regarding norm-referenced tests, there are also positive aspects of these tests that most professionals consider useful. Norm-referenced tests are statistically sound (Ornstein, 1993). Norm-referenced tests also tend to have high reliability. In contrast, curriculum based measurements tend to be locally made, the quality of the test items are often unknown and the item selection of these test questions may also be poor (Ornstein). Norm-referenced tests provide standards, which can be used throughout the country. When a child moves to a different school district knowing that the child was receiving "B's" might not be enough information to decide where to place the student. A "B" in one school district may be very different to a grade of "B" in another school district. The information from norm-referenced tests could be helpful in understanding the particular needs of the individual. There are many advantages to using norm-referenced tests despite their weaknesses. There are also ways to diminish their weaknesses.

Within the education field, the terms *norms* and *local norms* have come to have several different definitions. Therefore some definitions and explanations have been provided.

National norms- refers to an individual being compared to a large population from the nation. Intelligence tests are an example of a test that uses national norms. These tests publish a manual that contains the norms, usually by age, intended for the test examiner to compare the examinee (Sattler, 2001).

Local norms- refers to an individual being compared to a smaller population. Individuals may be familiar with the term *local norms* in relation to Curriculum-Based Measurement (CBM), which involves using standardized materials and comparing students to either other students within their class or school (Shinn & Bamonto, 1998). *Local norms* can also be developed for norm-referenced tests. This procedure would involve using a norm-referenced test, however, instead of comparing the individual to the *national norms*, the individual would be compared to the *local norms* (Stewart, 2005).

With regards to *local norms* there are different levels in which the norming process can be applied. The levels essentially refer to the group of students included in the norm sample. Thus the use of *classroom norms* compares the individual to others in their classroom, whereas *school norms* would allow for the individual to be compared to others within their school and *district norms* would allow for them to be compared the students within the district (Stewart, 2005).

Often when researchers and educators believe that the normative sample is not representative of the sample they wish to obtain information, they develop *local norms*. There are different reasons why a target population would differ from the norm sample such as economic or geographic, cultural or linguistic differences (Kamphus & Lozano, 1984).

Research on *Local Norms*:

Johnson, Taback, Escobar, Wilson, and Beitchman (1999) recognized that the Test of Adolescence/Adult Language-3 (TOAL-3) did not have a norm sample that was similar to their target population. Seventy percent of the individuals within the 18-25 age group for the TOAL-3 norms, had some post-secondary education. Johnson et al. were interested in young adults who had a history of speech and language problems. It was

hypothesized that students with a history of speech problems did not have the same amount of post-secondary education as the norm sample and thus the norm sample, was not a good representation of their target sample. The researchers explained that due to this highly educated norm group, even someone that fell within the normal range for speech and language problems might be identified as having a problem, In order to obtain information about the lower range of scores and to get a more representative sample these researchers developed local norms. Researchers found that these two groups were from different populations at a significance level of .01. In this study, education level affected how the construct of speech and language was measured.

Gronna, Jenkins, and Chin-Chance (1997) evaluated the use of the Stanford Achievement Test (Stanford 8) to assess academic performance in Hawaii. Although Hawaii uses the Stanford 8 for mandated assessment of students, no children in Hawaii were included in the norm group for either the national norms or the Pacific norms. The Psychological Corporation excluded students from Hawaii in their norm sample due to Haiwaiins comprising only 1 to 2 percent of the national population (J. Mayo, personal communication (as cited in Gronna, Jenkings, Chin-Chance, 1997)).

Gronna, Jenkins, and Chin-Chance (1997) explored the ethnicity of the public school population in Hawaii and found that there are 23% Caucasian, 4% Hispanic, .4% African-American, 21% Native Hawaiian and Part Hawaiian, 34 % Asian and 18% other. This ethnic make-up was different from the ethnicity found in the norm sample. For these reasons the researchers decided to develop local norms for the Stanford 8 to be used with Hawaii's public school population. In analyzing the data the researchers found that Hawaii had reading norms that were lower for each grade level that was tested.

However, the Hawaiian students out performed the national norms for math in three out of the four grade levels tested. Local norms provide an opportunity for a student to be compared to others with a similar ethnic background. School personnel would be able to identify at-risk students for their population and provide additional support where needed.

The Need for *Local Norms* for Native American populations:

Salois (1999) developed local norms for the Northern Cheyenne and Blackfeet Reservation. Native Americans are often included in standardization samples however; their tribal affiliation is not delineated. Each tribe has its own culture as well as differing socio-economic status. Past research on Native American children have shown that mean discrepancies vary between different tribes. The exposure to English language seems to affect IQ scores. Some tribes speak English as a primary language while others do not (McCullough et al., 1985).

Maclellan and Nellis (2003) examined the norm sample of the WISC-III. They observed that only 4% of the standardization sample was made of up of minority children. When trying to ascertain minority children's IQ, examiners often use nonverbal tests such as the Universal Nonverbal Intelligence Test (UNIT), the Test of Nonverbal Intelligence (TONI-III), or the Comprehensive Test of Nonverbal Intelligence (CTONI) (Bracken & McCallum, 1998; Brown, Sherbenou, & Johnson, 1997; Hammill, Pearson, & Widerholt, 1996). However these tests do not necessarily provide the same information as other cognitive measures, which lead examiners to use other cognitive measures that are not representative of the population with whom they are working. For example, research on Navajo children showed that their scores on the performance scale on the WISC-III and WISC-R are significantly higher than their verbal scale scores

(Naglieri, 1982; Tempest, 1998). Many Navajo children have limited English abilities, which results in the verbal scale measuring their English proficiency rather than their Verbal IQ. Examiners who understand the implications of using the verbal scale with minority children often do not report this information nor use it for decision-making. Instead they rely solely on the performance scale. McCallum, Bracken, and Wasserman (2001) caution examiners who use performance tests to assess non-English speakers, that minimal verbal skills are still needed since the directions for the nonverbal tests tend to be verbal. If the students do not understand the directions then their performance on these tests will be an underestimate of their abilities.

McLellan and Nellis (2003) analyzed the test scores of 175 Navajo children and established local norms on this population. They found that their sample and the norm sample had no significant difference between the means on the performance scale. However, when analyzing the difference of means between the Navajo children and the norm group for the verbal scale, the difference was statistically significant. The mean of the Navajo children fell one standard deviation below the norm group.

Factors that support the need for *Local Norms* with deaf populations:

Braden (2005) emphasizes that hearing loss and language acquisition affects deaf children's scores on norm-referenced tests. Most deaf children do not have early language models in life due to having hearing parents. This delay in language acquisition effects how their knowledge base develops (Braden).

It is assumed that an individual's score within the verbal domain of intelligence tests is a true reflection of his/her ability. With deaf individuals this assumption may not be true. Instead the score may reflect how much exposure they had to spoken language

and when they acquired their language skills. Research shows that verbally loaded tests are correlated with an individual's hearing loss (Braden, 1994).

A deaf person's hearing loss and language acquisition, are aspects that might affect his/her score on a norm-referenced test, another aspect is his/her culture. Currently many deaf people see themselves as belonging to a cultural minority (Padden, 1996). Padden (1996) discusses how Deaf people have their own schools, clubs, and churches. The reason that Deaf people have separate organizations is that their values tend to differ from the hearing culture. One of the biggest factors in discussing the existence of deaf culture is the fact that deaf people have their own language. Within the United States many deaf people use American Sign Language (ASL) to communicate. The fact that deaf people use a language other than English supports the view that there is a Deaf culture. With regards to assessment, an examiner will need to conceptualize deafness from a medical and a cultural/linguistic perspective to properly choose an assessment technique that will not be biased toward the population with which they are evaluating.

Research and ethical standards dictate that evaluators carefully choose assessment tools and methods when assessing deaf children. According to *Standards for Educational and Psychological Testing,* it is important to analyze how the target population may differ from the norm group (American Educational Research Association, 1999). According to Ysseldyke (2004), characteristics, which occur less than 1 or 2 percent of the time, are not necessarily found in the norm group. Since deafness is a low incident disability, deaf children are not likely to be part of the norm group of norm-referenced test. The exclusion from published norm groups as well as

other factors related to deafness may mean that the use of norm-referenced tests with deaf children could be problematic (Braden, 1994).

The utility of deaf norms:

For the WISC-R, *special norms* known as *deaf norms* were developed in attempt to improve the utility of the test for use with deaf children. Braden analyzed the need and utility of these norms. He was concerned about the validity of these norms. He claimed that if the test were valid for deaf children than deaf norms would not be needed (Braden, 1985). The validity was intact based on similar PIQ on both the WISC-R and WISC-III for hearing and deaf children (Braden, 1998). Braden believed that if deaf children do not have different cognitive abilities than hearing peers, then there was no need for deaf norms.

Although Braden argued that in general deaf norms were not needed, he did suggest that in certain situations they would be helpful (Braden, 1985). He explained that although deaf children produced similar results on PIQ compared to hearing children, the means and standard deviations of the deaf norms and the standardization sample are slightly different. This would therefore affect those children that are outliers within the deaf group. If hearing norms are used, more deaf children would be labeled severely mentally retarded and less deaf children would be labeled as gifted (Braden, 1985).

Weaknesses in using *Local Norms*:

Based on the varied results regarding deaf children and assessment, some researchers believe that developing *local norms* for deaf populations may increase the utility of information obtained by administered norm-referenced tests (Braden, 1985). However, *local norms* do have certain weaknesses. There are also fallacies that are used

to support the use of *local norms*. Braden (1994) elaborated how these certain fallacies of assessment are related to deafness. The egalitarian fallacy states that because there is a difference between deaf and hearing children's means, the test must be biased. A difference in groups could be attributed to something other than the test being biased (Braden, 1994). For example, when deaf students have lower verbal IQs than hearing student, it does not prove that the test is biased. Their lower IQs could be related to their English skills and when they acquired language.

The culture-bound fallacy is that by visual inspection one can determine which items would be culturally biased for a minority group. This fallacy is similar to that of the egalitarian fallacy in that the mere observance of a difference between groups proves that certain items are culturally biased (Braden, 1994).

The standardization fallacy promotes the belief that if a test does not include a minority group in the standardization sample, that the test is automatically biased toward that group. The corollary to this fallacy is if deaf individuals are included in the standardization sample then the test is not biased for a deaf individual. Those that believe in the effectiveness of deaf norms promote this fallacy. They believe that by comparing a hearing-impaired individual to a deaf individual from the deaf norms, test bias would be eliminated (Braden, 1994).

The development of *local norms* may seem as though it promotes these fallacies. However, this is not always the case. Although differences may appear between a target group and the norm group, those that are proponents of *local norms* do not assume that this means the test is biased towards this group. As mentioned before the deaf individual's delay in development of language can be attributed to lower scores. Lower

verbal scores then do not necessarily mean that the test is biased toward deaf people.

As discussed in the standardization fallacy, it is believed that when hearing-impaired individuals are included in the standardization sample the test is not biased for all hearing-impaired individuals. This fallacy is based on the belief that hearing-impaired individuals are of a homogenous group. In fact, hearing-impaired individuals are a very heterogeneous group. *Local norms* allow hearing-impaired individuals to be compared to others who have other similar characteristics to them, not only the characteristic of hearing loss.

In analyzing the utility of *deaf norms* Braden (1985) discussed the issue of validity. Much of the research with regards to assessing deaf children focus on the reliability of norm-referenced tests. In developing *local norms* the focus is on the lack of reliability for deaf individuals. However, validity is also an important component to consider when assessing deaf children.

Tests used with deaf:

Most of deaf assessment research focuses on noverbal tests (Maller and Immekus (in press). An example of a nonverbal test frequently used by individuals who assess deaf children is The Universal Nonverbal Intelligence Test (UNIT) (Bracken & McCallum, 1998). The norm sample for the UNIT did not include children with disabilities. The test developers did however study a small sample of hearing-impaired students (N=106). This study did show that the hearing-impaired individuals performed lower than the hearing individuals on the UNIT. However, the effect size for these differences was small (Braken & McCallum, 1998). Conversely, Krivitski, McIntosh and Finch (2004) found

that deaf children performed similarly to hearing children on the UNIT. Neither of the two groups performed significantly higher or lower on any specific subtest.

When assessing deaf children, researchers who use tests that are not nonverbal still tend to focus on the nonverbal or performance components of intelligence tests. The Differential Ability Scales DAS is an intelligence test that contains a nonverbal section known as the Special Nonverbal Composite (SNC). This composite is intended for those individuals who are not English proficient (Elliot, 1997). According to Ricco, Ross, Boan, Jemison, and Houston (1997) who analyzed a Specific Language Impairment (SLI), and a Deaf/Hearing Impaired (Deaf/HI) group, found that both these groups obtained significantly lower scores than the mean of the control group. It should be noted that there were only 14 individuals that made up the Deaf/HI group. Their sample contained students from both residential and mainstream schools. (Riccio, Ross, Boan, Jemison, & Houston, (1997))

Maller and Immekus (in press) also discuss that experts in deafness are not consulted when developing tests, which can affect content validity. Maller's focus with regards to deaf assessment is on the validity of the assessment tools. Although she believes reliability to be an important component of test measurement, she believes establishing the validity of the test is the first priority. Within her research she uses the Rasch Model and Item Response Theory (IRT) to support her claims. IRT is a statistical method that investigates the differential item functioning of a test. Differential item functioning is designed to explain whether specific items function differently in different groups (Sattler, 2001). Test items should differentiate higher functioning students from lower functioning students. Students who have a higher IQ score should have a higher

probability of success in answering more difficult questions than students with a lower IQ score.

Maller (1997) used the Rasch Model to explore the item difficulty for the WISC-III subtests; Picture Completion, Information, Similarities, Vocabulary, and Comprehension. The results of the statistics indicate that the response pattern for deaf children was dissimilar to that of hearing children. The item difficulty analysis also compared the deaf children's response pattern to the standardization sample and also found it to be dissimilar. This means that the WISC III is not measuring the same construct in deaf children. It is important to understand that if deaf children were simply less intelligent than their hearing peers then all items would be more difficult for deaf than hearing individuals.

Maller (2000) used IRT to investigate whether test items function differently in different groups on the Universal Nonverbal Intelligence Test (UNIT). When investigating Differential Item Functioning (DIF), Maller (2000) did not find a significant difference between her deaf population and the norm population. Currently IRT has not been used to investigate the validity of the WISC-IV and DAS for deaf children.

Thus far the discussion has focused on the importance of establishing the reliability and validity for tests that are used on deaf and hard-of-hearing children. The research appears to not distinguish between individuals who are deaf and between those who are hard-of-hearing. Research also tends to draw from one type of hearing impaired population rather than a representative sample of all individuals who are hearing impaired. Most of the research whether on reliability or validity has focused on deaf children, primarily from residential schools (Braden, Maller, & Paquin, 1993). When

developing *local norms*, investigators should investigate if there is a significant difference between means for deaf versus hard-of-hearing children. There might also be a significant difference between those that attend residential schools compared to those that attend mainstream schools. Usually deaf and hard-of-hearing children are grouped together, however, creating separate *local norms* for these two groups might be beneficial. Gathering information on *local norms* for assessment tools currently used, may help provide insightful information that can later be used for developing unbiased assessment tools for the deaf and hard-of-hearing children. *Local norms* help to address the social standards of the environment and reduce cultural bias (Canter, Lau, & House, 2002). This is particularly important within this study in which the hearing-impaired individuals differ from the deaf children who are typically studied and *local norms* developed.

The present study will evaluate the need for *local norms* with respect to the DAS, UNIT, and WISC-III. Due to the smallness in size of the special population, hard-of-hearing and deaf children were not separated into two separate groups.

Hypotheses:

1. Deaf and hard-of-hearing students will not differ from the standardization sample on any of the UNIT's subtests or composite scores.
2. Deaf and hard-of-hearing students will differ from the standardization sample on WISC-III and DAS subtests that require hearing and/or a linguistic ability.

Method

Participants

Participants in this study were 84 hearing impaired students. The data was collected over 13 years. The evaluator who assessed these students works for an organization that is a cooperative extension of ten suburban schools in Western New York. There were 35 males and 49 females who were assessed. Of the 84 students, 26 are classified as hard of hearing, 5 multiply disabled, and 53 deaf. Three of these students attended residential schools and 81 attended a mainstream school. Fifty-two of the students had early intervention services, 21 did not have any early intervention services, and 11 of the cases did not have information on whether they had early intervention services. Of the 84 students, 20 of the parents were hearing, 12 were deaf, 1 was hard of hearing, and parent hearing status was not indicated for 51 of the cases.

Procedure

The Differential Ability Scales (DAS) N= 49, the Universal Nonverbal Intelligence Test (UNIT) N= 49, the Wechsler Intelligence Scale for Children- Third Edition (WISC-III) N=14, were administered by a single examiner to determine the cognitive abilities of the population. The examiner is a nationally certified school psychologist, is knowledgeable in assessing deaf children, and can effectively communicate in American Sign Language (ASL). The data presented is archival data, in that the examiner assessed these students between the years of 1990 to the present (2003). The examiner determined at the time of the evaluation, which cognitive measures would be used on each individual. Therefore one individual was not measured using every test

Results

At the time of the evaluation students' raw scores were converted into standard scores based on stated test manual procedures. The standard scores were recorded for DAS clusters and general cognitive ability, the UNIT quotients and full scale IQ, and the WISC-III clusters and the full scale IQ. According to their respective test manuals the composite and full scale scores have a mean of 100 and a standard deviation of 15. The DAS subtest raw scores are converted T scores with a mean of 50 and a standard deviation of 10. The WISC-III subtest standard scores have a mean of 10 and a standard deviation of three.

Crocker and Algina (1986)'s method on how to develop local norms was used for this study. A one-sample t-test was run to determine if the deaf sample performed significantly different than the norm group population. For scores that contained a significance difference, norms were developed around the mean and standard deviation of the deaf population. The new norms were developed by calculating a z score and multiplying it by the standard deviation and adding it to the mean of the norm group. This procedure allows the examiner who works with this specific population to determine how a student functions compared to this specific population of which they are a part (See Appendix A)

Discussion

The first hypothesis that was developed based on the information from Krivitski, McIntosh and Finch's (2004) study, was that there would be no significant difference between the published means and the means for our population for the UNIT. This hypothesis was confirmed. It was found that the scores on the UNIT memory quotient,

symbolic quotient, nonsymbolic quotient, cube design, spatial memory, analogical reasoning, object memory, and mazes did not differ significantly from the published norms.

The second hypothesis was that any test that was biased towards hearing loss or ASL as a student's primary language would be significantly lower than the published norms. Deaf and hard-of-hearing score means for the DAS Verbal cluster and Word Definition scores were significantly lower from the published means. The Vocabulary scaled score on the WISC-III was also found to be significantly lower for the deaf sample. *Local norms* were developed for all of these variables (Table 1). As predicted by Riccio, Ross, Boan, Jemison, and Huston, (1997), testing deaf children's skills using tests that require verbal abilities produces scores that are significantly different from the published means. This current study found that the DAS verbal cluster as well as the word definitions subtest produced significantly different results.

In this current study the use of the special nonverbal composite was analyzed. In contrast, the sample of deaf and hard-of-hearing students (n=26) was not found to be significantly different from the norm sample. According to results of this current study special nonverbal composite of the DAS might be an appropriate assessment tool to use with deaf and hard-of-hearing children.

Braden, Maller, and Paquin (1993) have investigated differences between IQ scores for residential and mainstream schools. Thus it is important that more research focuses on deaf and hard-of-hearing individuals who attend mainstream schools since this population appears to be neglected from research although there are known differences between residential and mainstream students. In this current study, 81 out of 84 students

attended classes in mainstream settings and 26 out of 84 students were hard-of-hearing. This may mean that this population is different from the populations that are normally studied. More research should be conducted on hard-of-hearing students as well as students with a hearing disability within a mainstream setting.

There are limits to the usefulness of this current study. Generalizabilty to other populations is very limited due to this unique population of deaf and hard-of-hearing students from mainstream schools from the Western New York region. The results were also based on archival data. Thus the small sample size is a compilation of students who were tested over a ten-year period. There are some characteristics that may be different from the students within this current population who were tested ten years before compared to those who were tested more recently. The information is based on tests that are outdated. At the start of the ten-year period the norms for the test could be fairly new. Although some for some the test they are currently outdate. Those that know of the Flynn effect (Sattler, 2001) realize that students who took these tests more recently might have produced higher scores than those who were the same ability that were tested at the start of the ten-year period. Currently examiners are using the WISC-IV and a new version of the DAS is expected soon.

The current focus of deaf assessment appears to center on the validity of intelligence tests used with deaf students. The new versions of commonly used intelligence tests need to be investigated by IRT to establish their validity. Once it is established that these tests are valid for deaf students then the focus can return to creating *local norms*.

References

American Educational Research Association. (1999). *Standards for Educational and Psychological Testing.* Washington: DC. Author.

Bracken, B., & McCallum, R. (1998). *Universal Nonverbal Intelligence Test.* Itasca, IL: Riverside.

Braden, J. (1985). WISC-R Deaf Norms Reconsidered. *The Journal of School Psychology,23,* 375-382.

Braden, J. (1994). *Deafness, Deprivation, and IQ.* New York, NY: Plenum Press.

Braden, J. (1998). *Assessment of Hearing-impaired and Deaf Children with the WISC-III.* In: WISC--III clinical use and interpretation: Scientist-practitioner perspectives. Prifitera, Aurelio; Saklofske, Donald H.; San Diego, CA, US: Academic Press, Inc.

Braden, J. (2005). *Hard-of-hearing and Deaf Clients: Using the WISC-IV withCclients who Hard-of-hearing or Deaf.* In: WISC-IV clinical use and interpretation: Scientist-practitioner perspectives. Prifitera, Aurelio; Saklofske, Donald H.; Weiss, Lawrence G.; San Diego, CA, US: Elsevier Academic Press.

Braden, J., Kostrubala, C., & Reed, J. (1994). Why do Deaf Children Score Differently on Performance vs. Motor-Reduced Nonverbal Intelligence Tests? *Journal of Psychoeducational Assessment, 12,* 357-363.

Braden, J., Maller, S., Paquin, M. (1993). The Effects of Residential Versus Day Placement on the Performance IQS of Children with Hearing Impairment. *The Journal of Special Education,* 26, 423-433.

Brown, L., Sherbenou, R., & Johnson, S. (1997). *Test of Nonverbal Intelligence* (3rd Ed.). Austin, TX: Pro-Ed.

Canter, A., Lau, M., & House. (2002). Best Practices in developing local norms in behavioral assessment. In: A. Thomas & J. Grimes (Eds.) Best Practices in school psychology IV (pp. 1099-1114). Bethesda, MD: National Association of School Psychologists.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart and Winston.

Davis, J., Elfenbein, J., Schum, R., & Bentlerm R. (1986). Effects of Mild and Moderate Hearing Impairments on Language, Educational, and Psychosocial Behavior of Children. *Journal of Speech and Hearing Disorders, 51,* 53-62.

Elliot, C. *Differential Ability Scales.* (1997) San Antonio, TX: The Psychological Corporation.

Eckert, T. & Sharpiro, E. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review, 24,* 497-508.

Gronna, S., Jenkins, A., and Chin-Chance, S. (1997). Creating Local Norms to Evaluate Students in a Norm-Referenced Statewide Testing Program.

Hammill, D., Pearson, N, & Widerholt, J. (1996). *Comprehensive Test of Nonverbal Intelligence.* Austin, TX: Pro-Ed.

Johnson, C., Taback, N., Escobar, M., Wilson, B., & Beitchman, J. (1999) Local Norming of the Test of Adolescent/Adult Language-3 in the Ottawa Speech and Language Study. *Journal of Speech, Language & Hearing Research, 42,* 761-777.

Kamphus, R., & Lozano, R. (1984) Developing Local Norms for Individually Administered Tests. *School Psychology Review,13,* 491-498.

Krivitski, E., McIntosh, D., & Finch, H. (2004). Profile Analysis of Deaf Children Using the Universal Nonverbal Intelligence Test. *Journal of Psychoeducational Assessment, 22,* 338- 350.

Maclellan, M. & Nellis, L. (2003) Using the WISC-III with Navajo Children: A Need for Local Norms. *Journal of American Indian Education,42,* 50-60.

Maller, S. (1997) Deafness and WISC-III Item Difficulty: Invariance and Fit. *Journal of School Psychology,35,* 299-314.

Maller, S. (2000). Item invariance in four subtests of the Universal Nonverbal Intelligence Test (***UNIT***) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18, 240-254.*

Maller, S. & Immekus, J. (in press). *The psychometric properties of intelligence when used with deaf and hard of hearing individuals: Practices and recommendations.* In J.L. Mounty & D.S. Martin (Eds.), Assessing deaf and hard of hearing adults: Issues and practices. Washington, DC: Gallaudet University Press.

Markwardt, F. *Peabody Individual Achievement Test-Revised.* Circle Pines, MN: American Guidance Service.

McCallum, S., Bracken, B., & Wasserman, J. (2001). *Essentials of Nonverbal Assessment.* NY: John Wiley & Sons.

Naglieri, J. (1982) Does the WISC-R measure intelligence for non-English speaking children? *Psychology in the Schools, 19,* 478-479.

Ornstein, A., (1993, October). Norm-referenced and Criterion-referenced Tests: An Overview. *NASSP Bulletin*, 28-29.

Padden, C. (1996). *From Cultural to the Bicultural: The Modern Deaf Community.* In *Cultural and Language Diversity and the Deaf Experience* (pp. 79-98). Cambridge, UK: Cambridge Press.

Parasnis, I. (1996). *Cultural and Language Diversity and the Deaf Experience.* Cambridge, UK: Cambridge Press.

Riccio, C., Ross, C., Boan, C., Jemison, S., & Houston, F. (1997) Use of the Differential Ability Scales (DAS) Special Nonverbal Composite Among Young Children with Linguistic Differences. *Journal of Psychoeducational Assessment, 15,* 196-204.

Rodriguez, M. (1997, January) *Norming and Norm-Referenced Test Scores.* Paper presented at the Annual Meeting of the Southwest Educational Research Association, Austin, TX.

Sattler, J. (2001). *Assessment of Children: Cognitive Applications* (Fourth Edition). San Diego, CA: Publishers Incorporated.

Shinn, M., & Bamonto, S. (1998). Advanced Application of Curriculm-Based Measurement: "Big Ideas" and Avoiding Confusion. In: Shinn, M. (Ed). Advanced Application of Curriculum-Based Measurement. New York, NY: The Guilford Press.

Stewart, L. (2005). Developing and Using Local Norms: Guidelines for Educators. *National Association of School Psychologists.*

Tempest, P. (1998) Local Navajo Norms for the Weschler Intelligence Scale for Children- Third Edition. *Journal of American Indian Education, 37,* 18-30.

Wechsler, D. (1991). *Wechsler intelligence scale for children* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2001) *Wechsler Individual Achievement Test* (2nd ed.) San Antonio, TX: The Psychological Corporation.

Ysseldyke, S. (2004) *Assessment: in Special and Inclusive Education* (9th Edition) Boston, MA: Houghton Mifflin Company.

Table 1

Significant One-Sample T-tests

Variable	N	X	SD	t	p
DAS verbal cluster	49	92.49	17.695	-2.971	.005
Word Definitions	49	41.65	11.135	-5.247	.000
WISC-III vocab	14	7.79	3.215	-2.577	.023

Appendix A

This procedure will be completed every time you test a deaf or hard-of-hearing student:

Enter Data:

1. Assign the student a case number
2. Using SPSS for Windows enter his/her information in the data view (go to variable view to obtain information on how to code information about the student)

Analyze Data:

3. Go to the Analyze menu on the top of the screen
4. Pull down the descriptive statistics and go to the descriptives button
5. Enter the variables DAS verbal, WD DAS, and WISC-III voc. Press ok. (You should get an output that contains the mean and standard deviation for these variables)
6. Make sure the mean and standard deviation are easily accessible

Compute Local Norms:

7. Once you have entered a new case that contains information on the DAS verbal, WD DAS or WISC-III vocab you will need to follow this procedure.
8. Go to the transform button on the top of the screen, pull it down to the Compute button.
9. In the target variable title the variable "new DAS verbal".
10. In the numeric expression box use this formula:

((score – sample mean)/sample standard deviation)*norm standard deviation + normative mean

11. Here is example using the current information on DASverbal with a published normative mean of 100 and normative standard deviation of 15. The specific sample mean is X= 92.49 and a SD= 17.69.

((score- 92.49)/17.69)*15+ 100

12. Remember that you must use the normative mean and standard deviation from the manual. For example when entering a scaled score you would use the mean of 10 and SD of 3.

Once a year:

13. Run one-sample t-tests to analyze if other variables have become significant or if the variables that are currently significant are no longer significant
14. Go under Analyze to Compare Means to One-Sample T-tests
15. Click on all variables that use 100 as the means and enter this as the test variable
16. Run t-test, check significance
17. Do this same procedure for means of 50 and 10
18. To determine significance look at the significance column if the value is less than .05 it is significant
19. If significance is found *local norms* should be developed for these values, follow the instructions above.