[The following selection is taken from: Pollard, R. Q (2002). Ethical conduct in research involving deaf people. In V. A. Gutman (Ed.), *Ethics in mental health and deafness* (pp. 162-178). Washington, DC: Gallaudet University Press.]

Although this selection discusses "research data gathering tools", it is really about psychological testing. The information is just as relevant to clinical assessment situations as it is to research. Again, this is not the entire chapter, just the section on testing.]

Research Methods

One of the greatest ethical challenges in conducting mental health research with deaf and hard-of-hearing people is that addressed in guideline 2(b) above, which requires that research methods be appropriate in the cross-cultural setting. The risk of bias in data gathering and evaluation methods is very great. The aforementioned concerns regarding fund of information and literacy in the deaf and hard-of-hearing population are relevant here, as are sources of bias that may arise from the different sensory and, often, sociocultural experiences of deaf and hard-of-hearing individuals in contrast to hearing people. Rosen (1967) aptly demonstrated how misleading research data could be when personality tests designed for hearing persons are administered to deaf individuals. Yet, such evidence of bias does not necessarily mean that only psychological tests developed for, or normed with, deaf and hard-of-hearing individuals are appropriate for research. Judging the fairness and utility of psychological tests or other data gathering tools for use in cross-cultural research is a complex matter (American Educational Research Association, et al., 1999; Sandoval, et. al., 1998).

If deaf and hearing subject samples yield different but reliable results on a given research tool, this may or may not indicate a problem with the tool or the data gathering method. If the tool is allowing error to invade the data or is leading to bias suggestive of erroneous conclusions, then alternative methods, test revisions, or deaf or hard-of-hearing norms may be needed. An example would be the Facial Recognition Test (Butters & Albert, 1982), a neuropsychological tool where respondents are asked to identify pictures of famous people, including Golda Meir, Bob Hope, Fidel Castro, and Farah Fawcett. Fund of information limitations in deaf participant samples (e.g., unfamiliarity with entertainment and political figures due to lower literacy or the minimal captioning of television programs before the 1990s and the still infrequent captioning of news broadcasts), could lead to lower scores on this test and erroneous conclusions suggestive of neuropathology. On the other hand, the routine development of deaf population test norms could lead to problems if deaf/hearing data differences are real and not attributable to error. In such cases, renorming could obfuscate potentially important research findings. For example, if depressive symptoms are more prevalent in persons who lose their hearing later in life (Pollard, 1998), then development of special norms for late-deafened individuals on a test of depressive symptomatology might lead a researcher to overlook the significance of a (re-defined) "normal" degree of depressive symptoms in a late-deafened participant sample.

To judge the appropriateness of data collection tools in research involving deaf or hard-of-hearing participants, one must carefully evaluate five elements of the tool being considered. First, is whether its purpose is relevant to the question at hand. A measure of reading ability may be needed in a particular study but use of the Wide Range Achievement Test – 3 (Wilkinson,

1993) would be ill-advised when employing a deaf participant sample. This test gauges "reading" ability through a task involving word pronunciation. While word pronunciation and general reading ability may be sufficiently correlated for hearing participant samples, this assumption would not extend to deaf participant samples.

The second element to consider is the nature of the task <u>instructions</u>. Literacy issues, fund of information, and the adequacy of communication arrangements with deaf and hard-of-hearing participants are all potentially critical to the appropriate conveyance of instructions. Tests or other data gathering methods with long or complicated English-based instructions are particularly likely to present a challenge to the fair and complete conveyance of information to deaf research participants. Highly skilled interpreters or sign-fluent researchers are imperative in such situations, as well as the opportunity for participants to discuss the instructions at length until they are fully comprehended. Never should instructions be assumed to be self-explanatory, especially if hearing research participants would have been provided with verbal instructions. This error is commonly made when deaf individuals are given tasks that appear to be highly "visual," such as the performance subtests of the Wechsler Adult Intelligence Scale – 3 (Wechsler, 1997). Mimed, demonstrated, printed, written, "lip-read," or inadequately signed research task instructions seldom will substitute for the degree of task comprehension that hearing participants would experience in their preferred communication modality.

The third element to consider in judging the fairness of data gathering tools is the <u>nature of the task</u> that participants are asked to engage in, including the nature of any test items they must respond to. Test items and research tasks must be appropriate, not only to the purpose of the study and the participants' communication preferences and proficiencies, but also in relation to their sensory abilities (e.g., hearing loss), fund of information, socialization experiences, and cultural affiliation. Brauer, et al. (1998) discuss Messick's (1995) related concepts of <u>construct underrepresentation</u> and <u>construct-irrelevant variance</u> in the selection of test items or tasks used to assess deaf individuals. The Trail Making Test, Part B (Army Individual Test Battery, 1944; Spreen & Strauss, 1991), is a neuropsychological measure where respondents sequentially connect numbers and letters of the alphabet, but must alternate between numbers and letters (i.e., 1-A-2-B-3-C, etc.). Errors or tardiness in the connection order lead to speculations regarding neuropathology, but only because the test presumes respondents have an "overlearned" familiarity with the English alphabet. It is not unusual to find deaf individuals who do not have an overlearned degree of familiarity with the order of the English alphabet. Thus, even if the test directions are clearly conveyed, the nature of this task may lead to erroneous conclusions with deaf participant samples. Bias in the sensory or sociocultural content of test items is evident in the Minnesota Multiphasic Personality Inventory – 2 (Butcher, Dahlstrom, Graham, Tellegen & Kaemer, 1989), the most widely used personality test in the world. It contains items such as: "I think there's something wrong with my hearing"; "I find it hard to make talk when I meet new people"; and "I like poetry". The implication of "scored" responses to such items typically is psychopathology or personality deviance, not sensory or sociocultural differences in an otherwise normal individual.

The fourth element to evaluate in a data gathering tool is the <u>response modality</u> involved. Here, it is English-dependent response requirements that are most likely to lead to bias or error. For example, respondents to the Suicide Probability Scale (Cull & Gill, 1988) indicate how

frequently they experience certain thoughts, feelings, or behaviors associated with suicide risk. To each item on the test, they must indicate "none or a little of the time", "some of the time", a "good part of the time" (emphasis added), or "most or all of the time". Even if the test is appropriate to the purpose of the study and the test instructions and items are clearly understood, comprehension of the response condition, a "good part of the time" is critical to obtaining accurate data. "A good part" is an English idiom (which has nothing to do with the usual meaning of "good") that might not be readily understood by some deaf participants without further elaboration.

The fifth element to consider in evaluating a data gathering tool is the <u>scoring</u> method employed. Scoring methods must be appropriate in relation to each of the issues raised above (e.g., communication modality, sensory and sociocultural experiences) and, if normative based, must employ appropriate standardization samples. This may or may not imply the need for deaf or hard-of-hearing test norms, as noted earlier. The Symbol Digit Modalities Test (SDMT; Smith, 1982), a screening tool for neurological impairment, has no task or scoring bias that should require a special normative sample. The task content (rapidly matching digits and symbols according to a key), and the scoring (counting correct or incorrect symbol-digit matches) should be fair for hearing, deaf, or hard-of-hearing respondents, regardless of their preferred communication modality (<u>if</u> the instructions were adequately conveyed) or sociocultural characteristics, provided they are very familiar with the numerals 0 through 9. Special norms for the SDMT should not needed and, if developed, could obfuscate evidence of neuropathology that might exist in a deaf or hard-of-hearing research sample (e.g., if the etiology of hearing loss was associated with neuropathology, as might occur with prematurity, meningitis, anoxia, etc.).

The Rorschach inkblot test looks ideally suited for use with deaf candidates until the matter of scoring is examined. The Rorschach's purpose is to yield data relevant to personality and certain psychological disorders. Its content is visual and the nature of the task is to describe, in one's preferred language, the images perceived in the blots. Assuming the test's minimal instructions are conveyed appropriately to deaf respondents, and their answers are translated from sign to English by a skilled interpreter or bilingual researcher, it would seem that the Rorschach is an appropriate research tool based on the four criteria addressed above. However, the scoring methods usually employed with this test (Exner, 1993) are deeply embedded in styles and norms of English speech and could lead to substantial bias when applied to responses translated from ASL. For example, perceived movement is an important Rorschach scoring element. Even movement that is passive, such as a percept involving an animal that is merely hanging from a branch (vs. swinging) earns a "movement" score. Consider a deaf subject who perceives the image of an opossum in an inkblot. Assume further that the individual's fund of information is a bit limited, and that he or she has seen opossums, or pictures of them, but has not learned the name of the animal. In responding to the blot, the participant signs "animal," and then uses an ASL classifier that indicates the animal is hanging by its tail over a slim, cylindrical object, perhaps a tree branch. How should this response be translated? "Animal hanging?" "Animal hanging by its tail from a branch?" Or simply, "opossum?" How much movement of the classifier is needed to denote "hanging" from "swinging?" Each of these possible translation decisions would yield different Rorschach scores! Despite the considerable body of research on the use of the Rorschach with deaf individuals, scoring challenges with this test raise questions

as to its appropriateness with persons who respond in ASL or who are in any way not fluent in English. Hopefully, further research will continue to elucidate this issue.

In summary, when evaluating the fairness of a research tool, the key is to think through the data gathering method from start to finish, including the purpose of the tool and its goodness of fit to the research question, the plans for the conveyance of instructions, the item content or nature of the task, the response modality, and the scoring methods and norms. If, in any of these five areas, English knowledge, hearing loss, fund of information, or sensory or sociocultural aspects of life as a deaf or hard-of-hearing individual would play an undesirable role, then the test or data collection tool is suspect.