Strategies that can be used to determine bias in cognitive ability tests.

Author: Newturn Wikirefu . March 2021

According to Osterlind (1983), test bias is a systematic error in the measurement process which affects all measurements in the same way sometimes increasing and at other times decreasing it. In mathematical statistics, "bias" refers to a systematic under-or overestimation of a population parameter by a statistic based on samples drawn from the population. In psychometrics, "bias" refers to systematic errors in the *predictive validity* or the *construct validity* of test scores of individuals that are associated with the individual's group membership. Psychometric bias is a set of statistical attributes of a given test and two or more specified subpopulations. In layman's terms, the presence of test item bias means that two people who are at similar levels of the latent construct being measured but who belong to two different cultural, race or gender groups, respond differently to a particular question purporting to measure that construct, resulting in differences in the level of "performance" measured. When tests are labelled "biased", the accusations often have to do with the instruments chosen for a particular context, how these tests are administered or how the results are interpreted and/or used.

The concept of bias in intelligence tests can be conceptualised from different perspectives. According to Van de Vijver and Poortinga (1997), there are three forms of bias that need to be considered, namely construct bias, method bias and item bias. **Construct validity criteria of bias** can be considered under two main categories: external and internal. External refers to the correlations of test scores with other variables independent of the test itself. (Thus a test's predictive validity also may enhance its construct validity.) Internal refers to various quantifiable features of the test data themselves, such as reliability, item discriminability indices, item intercorrelations, and other item statistics, as well as to the factorial structure of the test.

Situational bias refers to influences in the test situation, but independent of the test itself, that may bias test scores. Examples are the race, age, and sex of the tester, the emotional atmosphere created in the testing situation, cooperativeness and motivation of the person taking the test, time pressure, time of day, and the tone and content of the test instructions.

Content or language Variety:

This type of bias refers to content or language or dialect that is offensive or biased to test takers from different backgrounds. Examples include content or language stereotypes of group members and overt or implied slurs or insults (based on gender, race and ethnicity, religion, age, native language, national origin, and sexual orientation) or choice of dialect or variety that is biased to test takers.

Group Performance:

This type of bias refers to the difference in performances and resulting outcomes by test takers from different group memberships. Group differences could occur among salient groups (e.g., gender, race and ethnicity, religion, age, native language, national origin, and sexual orientation) on test tasks and subtests.

***Standard Setting*:**

This type of bias refers to standard-setting in terms of the criterion measure and selection decisions and how these decisions affect different test-taking groups.

Construct bias [e.g., overlap in definitions of the construct across cultures, "differential appropriateness of behaviours associated with the construct in different cultures" and poor sampling of relevant behaviours associated with the construct.

Method bias [i.e., bias about the sample (e.g., samples are not matched in terms of all relevant aspects, which is nearly impossible to achieve), instrument (e.g., differential familiarity with the items), or administration (e.g., ambiguous directions, tester/interviewer/observer effects).

Item bias due to "poor item translation, ambiguities in the original item, low familiarity/appropriateness of the item content in certain cultures, or influence of culture specifics such as nuisance factors or connotations associated with the item wording"

Various strategies can be used to determine and guard against test bias. According to Owen (1998), ***construct comparability*** is the most fundamental issue, because it concerns the nature and essence of what is being measured. This can be assessed by, for instance, factor analysis and comparison of reliabilities for different groups. In the case of the Learning Potential Computerised Adaptive Test (LPCAT), a clear one-dimensional factor structure was shown for the LPCAT items for all groups concerned and comparable reliabilities for various subgroups were found (De Beer, 2000b).

Regression Analysis is yet another strategy that can be used to guard against the issue of test bias. If the factor structures are different for different populations, the instrument is not tapping into the same phenomena for the populations. Regression analysis can be applied to see if the tests make similar, and similarly accurate, predictions between the tests and a criterion measure. If, for example, regression slopes for a test or evaluation procedure and a criterion differ for different groups, test bias exists. Such studies require that we have fairly clear-cut criteria on which to judge the adequacy of predictors. Although some researchers (Kaplan & Saccuzzo, 1982) believe that slope bias for ethnic minority groups has rarely been demonstrated in empirical studies, we found convincing evidence for slope bias in the case of Asian Americans.

Under such circumstances, the instruments can be modified to enhance their validity or local norms can be established with different populations. Such efforts are important in that they provide a standard by which to compare different groups and yield insights into what aspects or items of a measure are crossculturally appropriate or inappropriate and what modifications may be necessary to strengthen the validity and to more accurately interpret test results.

Advocate for the integration of cross-cultural considerations in research, theory, and assessment practice. The use of measures developed in one culture and applied in another culture runs the risk of perpetuating an imposed emic in the assessment. That is, taking an emic (culturally specific) assessment scale and using it as if it were etic (universally applicable) in nature can be a serious problem.

Linear research is intended to examine the validity of an instrument. Whereas point research establishes that two cultural groups differ on a measure, linear research tries to establish whether the differences are real or an emic artefact of the measure.

Meta-analytic studies can also be used to eliminate bias in intelligence tests. A series of studies using different measures of a construct can be used with two or more culturally distinct groups, or different measures can be used in a single study. Linear research tries to establish whether the differences are real or an emic artefact of the measure. In parallel research, the task is to develop a means of conceptualizing the behavioural phenomena from the different cultures in question. A parallel design is essentially two linear approaches, each based upon its cultural viewpoint. The advantage of this design is that the framework or perspective from one cultural group is not imposed on another. In this way, similarities and differences of the constructor concept under investigation can be determined.

Standardisation and norming can also be used to eliminate bias in intelligence tests. This strategy seeks to see if the test or assessment instrument has been standardized and normed on the particular ethnic minority group. Increasingly, test developers are aware of the need to sample and validate tests and measures with different ethnic populations. For larger ethnic group populations, especially African Americans and Latino Americans, some measures have been standardized and normed.

Use multiple measures or multimethod procedures to see if tests provide convergent results. Before concluding, it is important to confirm findings from one instrument. This confirmation process should involve the administration of several different measures or different methods (e.g., behavioural ratings as well as self-reports) to see if the results are consistent.

Contextualisation is another strategy that can be used to eliminate test bias. This strategy seeks to understand the cultural background of the candidates, to place test results in a proper context. Ethnic minority groups exhibit significant heterogeneity and individual differences. Individual differences exist in 'country of origin, language is spoken and English proficiency, level of acculturation, ethnic identity, family structure, cultural values, history, etc. These differences have important implications for the ideal selection and interpretation of test results.

The use of ***assessment interpreters*** can also be used to eliminate bias in intelligence tests. Interpreters and their use in the assessment process have been questioned (Juarez, 1983; Langdon, 1988; Marcos, 1979). When poorly trained interpreters have been used, the results obtained in many instances may be worse than if no. Hernandez (1987) discussed reality factors related to the selection, training, and certification of Native Alaskan paraprofessionals and the cultural limitations that interfere with job performance. However, when well-trained interpreters are used and involved early in the assessment process, the result greatly improves both assessment and diagnosis of exceptional children. Many studies have reported positive results when using interpreters in very different cultural and linguistic settings.

(Cargo & Annahatak, 1985; Godwin, 1977; Marr, Natter, & Wilcox, 1980).

Conclusion

Flowing from the above discussion it is clear that bias in intelligence tests manifest itself in different ways and different strategies can be used to determine and guard against such biases as demonstrated in the above discussion.

References:

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

Abramson, T. The influence of examiner race on first-grade and kindergarten subjects' Peabody Picture Vocabulary Test scores. Journal of Educational Measurement, 1969, 6, 241-246.

Ali, F., & Costello, J. Modification of the Peabody Picture Vocabulary Test. Developmental Psychology, 1971, 5, 86-91.

Almy, M., & Associates. Logical Thinking in Second Grade. New York: Teachers College Press,

1970.

Almy, M.; Chittenden, E.; & Miller, P. Young Children's Thinking. New York: Teachers College Press, 1966.

Alpers, T. G., & Boring, E. G. Intelligence test scores of northern and southern white and Negro recruits in 1918. Journal of Abnormal and Social Psychology, 1944, 39, 471-474.

Altus, G. T. Some correlates on the Davis-Eells tests. Journal of Consulting Psychology, 1956, 20, 227-232.

American Association on Mental Deficiency. Adaptive Behavior Scale: Manual. Washington, D.C.: The Association, 1974.

Anastasi, A. Sources of bias in the prediction of job performance: Technical critiques. In L. A. Crooks (Ed.), An Investigation of Sources of Bias in the Prediction of Job Performance—A Six-year Study. Proceedings of an invitational conference. Princeton, N.J.: Educational Test ing Service, 1972.

Anastasi, A. Psychological Testing (4th ed.). New York: Macmillan, 1976.

Adebimpe, v. (1981). Overview: White norms and psychiatric diagnosis of black patients. American Journal of Psychiatry, 138, 279-285. American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., rev.).

https://thehumancapitalhub.com/articles/strategies-that-can-be-used-to-determine-bias-in-cognitive-ability-tests