

Bias and Cultural diversity in Psychological Assessment: An Empirical Review

Olusegun Emmanuel Afolabi

Department of Educational Foundation
Faculty of Education University of Botswana

Abstract: A century of research evidence on psychological assessment shows that scientist have made several efforts to advance a “culture free” tests (Jensen, 1980). Similarly, research also demonstrates that only a few numbers of issues in psychology research divide researchers and the general public as the use of standardized assessments with diverse culture. To illuminate these concerns and possibilities in a concrete context, the article systematically analyse the history of psychological assessment and explains the application of psychometric and socio-cultural framework for psychological tests. This article uses empirical evidence to analyses cultural bias in psychological tests and explores various approaches that describes and examine bias in psychological assessment. The paper also explores (1)the taxonomy of bias and equivalence in psychological testing,(2) identifies issues surrounding test bias, (3)explain sources of bias (4) evaluate how culture influences psychological assessment of diverse groups and last but not the least,(5)examine the inference of bias controversy and recommend various processes that remove bias in psychological assessment. Finally, findings reveal that psychological test performance on different cultural group shows different outcomes.

[Olusegun Emmanuel Afolabi. **Bias and Cultural diversity in Psychological Assessment: An Empirical Review.** Researcher 2014;6(2):42-53]. (ISSN: 1553-9865). <http://www.sciencepub.net/researcher>. 7

Keyword: Bias, Cultural diversity, Method bias, Psychological assessment

Introduction

All through the history of psychology, no problem has been so tenacious or as impervious to resolution as that of cultural bias in psychological test. Scientific debate since early nineteen century particularly, on intelligence has faced unprecedented scrutiny for cultural biases in different forms. Indeed, the history of intelligent test demonstrates that the problems of bias (or its potential) in assessment continue to be the source of recurring, characteristically intense and social controversy (e.g., see Reynolds & Brown, 1984). This evidences further support the notion that psychology as a discipline cannot achieve the inevitability and precision of the physical sciences, except it leans on research and measurement. This prompts the need to introduce series of psychological assessments and measurements to psychology research. Research shows that possibly no other assessment problem is as intense, contentious, and often discussed as that of bias in psychological test. This confirmed that only a few numbers of issues in psychology research divide clinicians and the general public as the use of standardized assessments with diverse culture, and this continues to be a major discuss in psychological research and assessment till date. Thus, the dominant subject concerning clients, parents, and clinicians are the long-standing consequences of variances in means test outcomes among ethnic group i.e. Blacks, Hispanics, American Indians, and Asian Americans, to mention a few.

Interestingly, a century of research evidence shows that scientist have made several efforts to advance a “culture free” tests (Jensen, 1980). The history of psychological assessment attested to this statement which further shows that efforts have been made in the past to create “culture-free” methods (Anastasi 1988; and Cattell 1940). While it is argued that the consequence of culture can be eliminated or measured in a test when spoken items were removed, and non-verbal items were used, research still reveals that this turned out to be wide of the mark in psychological assessment. Besides, literature establishes that the use of widespread diversity of cultural groups across the world occasionally detects greater group variability in performance and additional non-verbal assessments than the way it recorded in oral assessments (Anastasi 1988; Irvine & Berry 1988). These rationalization corroborate the existing fact that ethnic/cultural subgroup varied in standardized assessments (Herrnstein & Murray 1994). But reason for these differences remains a controversy till date. Based on this, psychology research is totally incomplete without proper application of psychological test.

Besides, more than any other professionals ‘psychologists are the main user of psychological tests in research. This process includes: intelligent test, aptitude test, personality test, cognitive test and mental test to mention a few. Although the debate about how tests influences the judgments of test administrator and the prospects of the testee continue to generate more

attention till date, however, the argument still remains, how can we eliminate bias in psychological research? For example, “an intelligence test is an unbiased, insignificant instrument until somebody allocates meaning to its outcomes. However, as soon as meaning is attached to somebody's score, the person will experience different consequences stretching from artificial to natural lifetime changing, which can be reasonable or biased, supportive or injurious, suitable or ill-advised subject to the connotation attached to the test score”. Similarly, a body of research on neuropsychological test shows that the procedures failed to identify an appropriate analytical accurateness when administered to non-Caucasian cultured natural English-speaking, as well as those from middle and upper socioeconomic status (Ardila, et al, 2002; Boone, et al, 2007; Brickman, et al, 2006; Loewenstein, et al, 1994; Manly 2005). Also, research continuously makes known that not only is this generates controversy, but that, psychological tests amongst ethnically and linguistically varied people is seriously susceptible to hypothetical and experimental basis in neuropsychological practice.

The purpose of the paper

This essay analyses the empirical evidence on cultural bias in psychological test and exploring various approaches that describes and examines bias in psychological assessment. The essay begins with brief overviews of history of psychological assessment and then explains the application of psychometric and socio-cultural framework to psychological tests. Also, the paper look at the taxonomy of bias and equivalence in psychology research and identifies issues surrounding test bias, sources of bias and how culture influences psychological assessment of diverse groups. Lastly, the paper examine the inference of bias controversy and recommend various processes that remove bias from psychological assessment.

Methodology

This paper employs and reviews literatures in order to analyze and assesses new empirical evidence that explain the influence of bias and cultural diversity on psychological assessment. The literature review process is carried out by searching for the key words, collating and reviewing relevant articles, books, journals, and meta-analysis about bias and cultural diversity in psychological test using the following online data base to seek for reference, choose relevant literature, investigate studies i.e. ERIC, PsychInfo, EBSCO host. This process reported about 4500 articles, journals, technical reports and paper presentation and book chapters covering more than 30 years period. Based on abstracts from the 4500 search reports articles cum journals, the search was lessened

to quite a few studies that are relevant to the research topic. Hence to achieve the aims of this paper, the content of the remaining several hundred journals or articles were scrutinised and only those that reported empirical findings were used. To confirm and verify references, manual searches of relevant journals and articles related to the topic are performed.

Literature Review

Historical Perspectives

Psychological test in its current arrangement originated a little more than a century ago in a research laboratory of sensory discernment, motor skills, and reaction time. Research shows that Francis Galton (1822–1911) designed the first series of tests, and odd variety of sensory and motor methods. Besides, James McKeen Cattell (1860–1944) an American psychologist having worked with Galton announced the recent testing outline in his model paper entitled “Mental Tests and Measurements.” in 1890. The paper indicates that the outcomes of the outline would be of huge systematic importance in determining the reliability of intellectual developments, their interdependence, and their difference under diverse situations. It also explains that the assessments will be thought-provoking, and, possibly, valuable with respect to teaching, way of life or sign of illness. In addition, the logical and practical worth of such assessments in a live setting would be substantially augmented if an identical scheme is embraced with resolutions made at various periods and situation is likened and shared (Cattell 1890). Based on this premise, the assumption that “possibly” the assessment would be helpful in “teaching, method of life or sign of illness” remains far-sighted understatement of all time.

Having raised more argument than any other issue in psychology research, bias in assessment continues to generate debates among scholars. This issue has being at the forefront since nineteen century when the first intelligence scale was developed and Stern presented measures to assess intelligence (Binet & Simon 1916, 1973; Stern 1914). Although this skirmish is not restricted to only intellectual aptitude tests; nonetheless, the debate about is influence continues to lure the attention of the general public till date. While numerous scholars and researchers have come up with publications that address cultural bias in test, the fact is that the issue remain contentious (Gould 1996; Herrnstein & Murray 1994). Moreover, the controversy surrounding intelligence tests attracts many court cases, incites government regulation, and brings flay from general media (Brown, et al, 1999; Reynolds 2000a).

Jensen (1984a) maintains that the conduct of test bias prior to the 1970s was disconnected, haphazard,

and theoretically muddled. This shows that perfect and commonly agreed-upon connotations of bias were not there in assessment. For example, the psychometrically strong procedure that empirically identifies test bias in psychological research are lacking, therefore making the outcome of the test to be biased and uneven. This prompts up cautiously thought-out justification and numerical procedure that psychometric had long advanced in the areas like consistency, validity, and item selection (Jensen (1984a: 507). Interestingly, the commencement of the early seventy, witnesses an enormous bulk of unbiased and experimental studies of multifaceted problems such as development and usage of standardized psychological assessments with American-born, English-speaking subclasses (e.g., see Berk, 1982). However, the general fact and age-old claim in psychological tests is that means variances are as a result of "cultural bias" in assessment. Jensen (1980) systematically reviewed the experimental studies on test prejudice and concludes that test predisposition is the most shared gathering opinion of critics. Besides, Jensen (1980) also concludes that standardized ability tests forecast correspondingly for English-speaking majority, American-born, and minority subclasses, and at the same time measures related concepts.

According to Gregory (2004) no exercise in contemporary research is more attacked like psychological tests and that analysts maintain an exceptional and often passionate criticism of aptitude testing in particular to date. This argument is rooted in perceived variances of typical intelligence scores particularly, among several ethnic clusters (Blacks) and racial clusters (immigrants) in early nineteen century (Cole & Zieky 2001). Also research establishes that African Americans score 15 points smaller than white colleagues on traditional aptitude assessments with extraordinary language/spoken and traditional loadings (Flanagan & Ortiz 2001). Therefore, conclusion on disparity group test score performance in aptitude test continually intensifies debate on test bias (Gregory, 2004).

Framework for Bias Detection

While a significant numbers of literatures identify various methods to bias detection (Berk 1982, Marascuilo & Slaughter 1981), scholars like McCauley and Colberg used analysis of variance to described and explained cultural interaction as a sign of bias in item. A test is biased if an item shows large number of cultural interaction and also transferable when it is zero. Most available literature on bias in psychological tests shows that unevenness occurs in the following situation: where there is deficiency in test administration, poor rendition of items, and tapping of personality in groups. Although this is

improbable in practice, nonetheless, it is conceivable that all test items are evenly biased and favour a particular group. This is wholly perceived as a major consequence for culture, and it makes item by cultural interaction insignificant. These points to the fact that, assessments like this interface indicate inadequacy in bias verification.

Psychometric framework.

Though the issue of pre-conceived notion remains valid subjects that analyses psychometric concepts, it is well reported that a test is not reliable if it shows inaccuracies in context when apply to recognizable subpopulations. The tasks to test objectivity were developed between the late 60's and early 70's as a consequence of a number of issues coming together. Some of these issues include: evolving anticipation of fairness in research outcomes; the emerging opinion that African-American beliefs are the same as that of leading white values; and the ensuing authentication of methods that accentuated its characteristic potentials. However, the most significant factor surrounding this progress is linked to the advent of black psychology. This development measures and analyses Afro-Americans principle (Hilliard 1995). Moreover, black psychology believes that appraisal need to be articulated through ethnic material acquainted to the assessment participant, and that a test must give consideration to histories of domination, as well as emotional persecution. Undoubtedly, such mind-set may possibly not depend on apparatus established in the old model. Based on this resolves, most consideration for assessment is directed at psychological assessments and traditional predispositions such as intelligence assessments or personality inventories. Nevertheless this without doubt spilled over into accomplishment testing.

Socio-cultural framework

While substitute opinion of the psychometric method sees assessments as a portion of the social marvels of public education, it's also established as an evidence of larger social and ethnic problems. Frisby (1998) enunciates three probable expert methods that explain culture: i.e. the practitioner-clinician, socially conscious advocate and theorist-researcher. This notion proffers an understanding and superior answer to questions of prejudice in testing as psychometric examination failed to place our effort in a social context. Nevertheless, because of poor knowledge of the cultural context, it is difficult as researchers to solve the issue of bias in testing. Although research acknowledged the fact that psychometric studies notice the presence of bias in tests, most studies shows that it is very hard to clarify it. Nevertheless, theorist-researcher tries to comprehend the causes of

performance through unbiased progress and appraisal of the concept by investigating the interactions between smaller group affiliation and educational achievement as steered mainly by empirical data. Despite the fact that it is evident in their approach that they are probing culture, results pointed out that they find it very hard and therefore observes assessment data only.

The research on test bias starts with the set of courses then focuses on the methods of measuring information and abilities. Test data helps describing the problems, nevertheless the objective is to cultivate an improved philosophy that explain what and by what means does students learn and in what manner does this influenced by cultural experience. Thus, closely tangled is the question, "What must be done to make sure that all learners study?" To answer this, effort should be directed towards uses of assessment tools that are consistent with both the cultural experience and the anticipated result. It is also imperative that practitioner-clinician must get close to the assessment director's role by searching for consistent knowledge to guide practice. This is logically determines by theory, and focuses on practical answers to daily difficulties. This illustration appeals to both psychometric and socio-cultural agendas and therefore, makes assessment an instrument toward pupil education rather than an end in itself.

However, for dependable awareness that guide training, a respectable practitioner-clinician must see to it that experimental outcomes are construed within a situation that contains the culture of the learner and the social moulds of the assessment. This is based on the beliefs that a learner needs to obtain irrespective of their cultural inheritance. Furthermore, socially perceptive supporter try to defend rights that historically-excludes groups. As similar to the theorist-researcher, the campaigner's probe the bigger culture and the learner's place within it. They also changes to what learners need to be acquainted with and the manner of measurement. Besides, the advocate shows little signs of worries about fairness or model development compares to altering significant injustices of domination and segregation. As a result of this notion, experts incline to ignore advocates for not knowledgeable in theory, but greatly obligated to their unyielding voice.

Test Bias Controversy: Definitions and Taxonom

Historically, psychology reveals quite a few examples of far-reaching overviews about variation in aptitudes and characters of ethnic people that were founded on psychometric poor processes. However, to avert creating far-reaching declaration which in the end disfavour to the subject, a lack of bias (i.e., equivalence) ought to be established as an alternative

for assuming (Poortinga & Malpass 1986). The review of literature establishes that assessments are frequently regarded as prejudiced against Black and other racially and linguistically varied groups. This is counter to students from lower socioeconomic group as it favours both middle class and white learners. According to Gregory (2004: 242) test bias is defines as "unbiased numerical manifestations that observe the patterns of assessment scores for appropriate subpopulations". Gregory (2004) continues by conclude that agreement occurs on the numerical standards which refers to an assessment as subjective.

Bias exists when an assessment score has connotations with a related, definable subclass of test participants and varied in connotations or effects for the remnants of test participants. Hence, bias is the degree of difference in validity of an assessment score for a definable and appropriate subclass of assessment participants (Cole & Moss 1998, cited in Gregory 2004: 242). However, from a social standards perspective, when a test is prejudiced, the worry is synonymous to renunciation of prospect and the improper undesirable assumption. Also, there are other standings that are germane to debates concerning testing CLD groups. Research shows that while a test is not subjective in principle, it might be biased (see Cole & Zieky 2001). Hence, assessment objectivity is basically relates to social effects of test outcomes (Gregory 2004, Hunter & Schmidt 1979).

Orientation to Issue of Test Bias

Jensen (1980) postulates that a suitable method for study test prejudice can only continue if shared, but misleading suppositions is recognized and put to rest. The shared misconceptions can either be (a) classless misleading notion, i.e. all human subclasses are alike or the same in personalities measured by assessments; (b) culture-bound misconception, i.e. assessment items is recognized or classified according to their "culture-loadedness" from unplanned assessment or bias decision; and (c) standardization error, i.e. an assessment is unavoidably subjective if apply to several populace than those comprised in big figures in the standardization specimen. Interestingly, Jensen (1980) maintains that assessment unfairness and objectivity are distinguishable objects and that impartiality is an ethical, lawful, and/or metaphysical matter makes sensible individuals legally differ. Thus, assessment bias is an empirically grounded numerical subject that applies to the psychometric properties of a particular assessment of two or more certain subpopulations (Jensen 1980: 375).

Types of Bias

To promote the assessment of bias the following classification that prejudice is related to disparity in

performance among or between groups is valuable. This brings about the following questions: why differences in performance among cluster (Black or White, female or male, high or low income) and why is this consistent? Efforts to justify the differences in test performance target the following factors: personal qualities of examinees, the testing situation, and nature of the assessment (Scheuneman 1985). Thus, this paper will examine the following classification of bias:

Construct Bias

Research indicates that bias occur when a construct measure is not matching an ethnic clusters or when an assessment quantify different theoretical concepts or characters for one cluster of people than another. This predisposition is also applicable when an assessment analyses similar attribute for a particular groups but shows contradictory degrees of precision. A decent case of construct validity is western aptitude assessments. Research establishes that in most cases general aptitude tests emphases on cognitive, learned knowledge, and recall and social features of acumen are frequently less accentuated. Besides, practical indication shows that the latter characteristics are more evident in non-Western situations (Super 1983). Review of literature on intelligent shows that psychology research is not fair to this particular area of study, i.e. learning. Statistics concerning factor arrangement are repeatedly indicates here as research proves that biased assessment demonstrate different element structures across subclasses. This indicates a lesser mark of resemblance for factor arrangement and also ranks or item exertion across clusters (Sattler 1992).

Also, construct bias is reported in a study on filial piety (i.e., conducts relates to a decent daughter or son (Ho, 1996). Contrary to Western cultures, Chinese children show extra and diverse duties to their blood relations. This dissimilarity may possibly be as a result of learning and income. Kagitcibasi (1996) confirms that in Turkey “assist with domestic responsibilities” are very rare in the life of parents with higher learning. Also, the importance of youngsters as old age sanctuary for their care givers declines with income level. So, an assessment of filial piety across ethnic populaces is vulnerable to construct bias.

Method bias

Reynolds (1998) defines method bias as a situation that make an item to a certain extent more hard for members of a particular group than the other if the overall competence level of the groups is constant and no sensible theoretical justification occurs to shed light on group variances in question” (Gregory 2004). For instance, questions like, “what

similarity exists between football and soccer?” a student or group that has little experience about how to play soccer will be at disadvantage. Lack of experience and knowledge about the game will put them at a disadvantage. Reynolds (1998) gives three occurrences to explain content bias: 1) Any objects that probe for data that lesser group of people do not have the same chance to acquire; when the scoring of the subscale is unsuitable because the assessment writer/inventor had subjectively categorical on the single right response and the minority groups are inaptly punished for stated responses that is correct according to their culture. Therefore, phrasing the questions is unaccustomed, and minority groups who know the answer will not respond since they do not have fair knowledge of the question(s) and are unaccustomed to the format of the assessment.

This process also denotes the difficulties originating from the instrument (instrument bias), and this is illustrated as stimulus awareness. Besides, Derogowski and Serpell (1971) studied Scottish and Zambian children in order to explain the miniature replicas of animals and motor vehicles, and different situation where they sort out pictures of these replicas. Although no culturally diverse were establish for the real replicas nevertheless, the Scottish children achieves better than the Zambian children when pictures are highly organized. This result shows that response techniques buoy up method bias. Serpell (1979) requests both Zambian and British youngsters to replicate design by paper-and-pencil, plasticine, shape of hand positions, and iron wire (creating replicas with iron wire is a common activity amongst Zambian boys). The British’s children recorded considerably better results in paper-and-pencil process while the Zambian’s children recorded better in the use of iron wires.

According to Hui and Triandis (1989) Hispanics are more inclined toward selecting excesses on a five-point assessment scale than White Americans. Nevertheless, this is absent when a 10-point scale was used. Moreover, administration bias occurs when communication glitches occur between interviewers and interviewees particularly, with diverse cultural and languages experiences (Gass & Varone 1991) i.e. inadequate awareness of assessment language by the interviewees’, Also, unfitting manners of address or traditional norm desecrations by the assessor (Goodwin & Lee 1994) can affect the gathering of suitable data.

Item Selection Bias

Research indicates several methods that help to identify bias in item selection (Holland & Wainer, 1993; VandeVijver & Leung 1997a, b). The common technique mostly used is the Mantel-Haenszel statistic

(Holland & Thayer, 1988; Klieme & Stumpf, 1991). This is practice when analyse bias in dichotomously recorded objects that is common in psychological assessments. Bias occurs where the items and the chosen tasks is founded on knowledge capabilities and linguistic of the main group, This is linked to content validity, and directly target the suitability of separate items. Statistically, the general assessment is not bias although some might have bias content. This explains why an item is included in onetest and not in other. As opposite to other method, item bias connotes misrepresentations that happen at item level. This has a diverse mental connotation across cultures, i.e. subjects' reactions in some group are somewhat cause by social appeal of item in a self-report inventory.

Similarly, the contrast of entire assessment marks between cultures is null and void when this item is incorporated. Besides, it is reported that item prejudice is given substantial consideration in the literature as most research on bias concentrates more on exploring and assessing statistical measures as a means of detecting item bias. Based on the statistical–methodological view, an item is influenced if diverse groups of people operationalized as the overall mark do not possess similar anticipated score (Shepard et al., 1981). On the other hand, individuals from diverse cultural groups with equivalent overall marks (i.e., individuals from dissimilar ethnic clusters who are similarly gifted, or nervous on what is measured), an impartial item ought to be similarly hard (or eye-catching). This specifies equivalent mean scores that transverse cultural group and therefore, any differences in scores signify bias.

Bias in Predictive Validity.

Bias exists when the conclusion from the assessment score is not done with the minimum practicable random mistake or when there is continuous mistake in interpret or predicts a task affiliated to a specific group. The main problem is: “Can the assessment scores precisely envisage learner’s performance in a given future assignment?” Research often assumes that a high aptitude score will lead to achievement in school and career. The opposite scholars worries is based on the fact that aptitude tests assumed much influence, and when learners or groups achieve low scores in aptitude test, there is a likelihood of them being deny the chance to use a program or facility due to negative perception. Besides, an assessment is well thought-out as “neutral if the outcomes for all related groups form one regression line. A neutral assessment foresees performance correspondingly for all clusters, even though they do not have the same mean” (Gregory 2004). Of significance, in correct forecast occasionally mirrors unpredictable dimension of the individual

being foretold rather than partiality in the assessment used for the prediction

A number of researchers address predictive bias in personality assessments (Moran 1990; Monnot, et al, 2009). Similarly, Monnot et al. (2009) confirms “modest” form of extrapolative bias across many measures. This study establish that “[t]hese prejudices point to both above- and below forecast of mental illnesses among African Americans on measures that signifies differential precision for the MMPI-2 in forecasting diagnostic status among group of inpatients male veteran looking for drugs misuse treatment” (Monnot et al :145). In contrast, no sign of over postulation of diagnosis is established for Caucasians in the assessment scores. In contrast to content and constructs’ validity, this result proposes light sign of prejudice, i.e. differential or single-group validity. Therefore, variations are not common and where it is notice, they take slight over predictions for lesser-scoring groups like those deprived, low-socioeconomic status, and tribal minority examinees. The over forecasts are improbable justifies for negative situation and analysis of those groups. Moreover, in exceptional scale, the slight dissimilarities establish might be a reflection and not the major causes of far-reaching social disparities affecting members of the social group. Therefore, origins of such difficulties like job discernment and poor socioeconomic status lie mostly outside the assessment setting.

Sources of bias

Past and present literature has done little to explain the probable causes of bias in testing. As we all believe that, bias is an issue in tests that works to swell score variances between two populace groups, it is imperative to state that analysing the sources of bias is vital if we are, to tackle the impacts that is presence has on research outcome and at the same time proffer effective solution to its occurrence. Recent theory and analysis of bias has come up with broader perspectives than those recommended by previous research, which lean towards other bias issues. Based on this viewpoint, questions about origins of bias can be more easily expressed and addressed in test. Also, most research literatures points out that the causes of bias in cross-cultural testing are multifarious, and therefore, difficult to give a thorough overview of the concept (Van de Vijver & Leung 1997b). Bias in construct validity takes place when there is limited overlay in the meanings of the concept across cultures. Western cultures ‘notions of aptitude are far-reaching and typically contain features such as social skills. This predominantly adds to the academic advancement in aptitude based on Western concepts (Serpell 1993; & Super 1983). Yang and Bond (1990) also carried out a

study on a group of Taiwanese subjects and reported that out of the five Chinese factors recognized; four of them are correspond to the American factors. This is corroborated by Cheung et al. (1996) on native Chinese behaviour measurements like “face” and “harmony”.

A Similar study by Church (1987) establishes a native behaviour concept in the Filipino culture. Also, Chinese Culture Connection (1987) came up with an assessment survey that is grounded on Chinese ideals and run in 22 nations. Besides, Hofstede (1980) recognises three of these factors based on Western standard. Nonetheless, factor such as Confucian Work Dynamism is connects to economic growth and isnot embracing by Western culture. Also, Hoshmand and Ho (1995) emphasises the significance of social characteristics in Chinese behaviour compare to the ideas of “independent self” in that dominate Western culture. This notion is generally embraced by several scholars (Bochner 1994; Paranjpe 1995; & Sampson 1988).

Research also establishes that administration bias is triggered by changes in the ecological administration circumstances, i.e. physical, technical, or social. This shows that disparity in the physical existence of dimension causes considerable cultural variances non-target variables like the matters’ level of inquisitiveness (triggered by the originality of the condition) or readiness to self-disclose. A related study on psychological tests by Tanzer et al., (1995) shows that participants in a location expresses the sign of coldness due to the air-conditioned in the assessment room that is beyond the researchers’ regulation. Moreover, research also reported different instances of social ecological situation such as individual vs. group administration; extent of space between testees (in group testing); or classroom size (in academic settings).

Furthermore, administration bias is also linked to communication among respondent and the interviewer. Therefore, problems associate with language remain the major source of bias. This is because most tests are conducted in second or third language of either evaluators or respondents (Gass & Varone 1991). Research also reveals that miscommunication arises from ethnocentric clarifications (e.g., Banks et al., 1991; Barna 1991; & Cohen 1987). Although research examines these worries nevertheless, little outcomes are obtainable for categorizing long-term social effects of testing. That is why Reynolds et al., (1999) maintain that difficulties are broadly pertinent to assessment than to ethnic problems. They also establish that it is pertinent to treat special education as a problem with education rather than assessment.

Effects and Implication of the Test Bias

Research has come up with approaches that detect and addresses different forms of bias. These plans tacitly shoulder that prejudice is an exasperation factor that ought to be evaded. As a result, research is committed to methods that allow the decrease or even removal of bias. This argument leads to an increasingly refined body of research that confirms that higher statistical test is generally impartial. In addition, research on neuropsychological assessments specifies that results are not definitive which indicates a slight bias in testing. Similarly, research probing psychophysiological methods to aptitude continues to produce results that explain the relationship between beliefs and mental functioning. Besides, researchers such as Verney, et al, (2005) confirms that the methods of managing information are applicable to Caucasian American students’ performance as oppose to Mexican American students. This finding further confirms differential validity in prediction. Research justifies that both discussions disregard these results and highlights as alternative mean variance among tribal groups (Reynolds 2000b). Moreover, publishers have come up with a nonverbal test that is recognized as: culture-reduced methods of ability; specialists construe scores, i.e to reduce the effect of supposed bias; and, last but not the least, an assessments to remove group variances. These reviews have adverse long-term effect on minority group members, and therefore, avoid the study of any prejudice that might be found.

Furthermore, other assessments procedure include: language and communication (Mpofu & Ortiz 2009). Moreover, the means of identifying bias in a nonverbal situation is similar to those uses in testing verbal content. Therefore, evidence of bias on nonverbal processes include: Comprehensive Test of Nonverbal Intelligence (CTONI; Hammill, Pearson, & Wiederholt 1997); Leiter International Performance Scale- R (LIPS-R; Roid & Miller 1997); and Universal Nonverbal Intelligence Test (UNIT; Bracken & McCallum, 1998) are assessed in seminal texts (Maller 2003). Dissimilarities in behaviour among ethnic and racial groups are abridged on nonverbal procedures. The consequence of these impacts is differs and depend on whether the bias elucidation is precise or improper if at all is acknowledged. Therefore, improper bias elucidation leads to adapted tests that are not mirror significant precise data. This report gives unsuitable data that unevenly groups had performed equally. Thus, Scholars ignorant or heedless of such disparities leads study into the reasons of these differences.

Discussion

Many methods is used to explain construct and method bias in psychological tests. The prominent among them is the use of informers that is verse and knowledgeable in native language and beliefs. Broer (1996) uses committee method to advance the first form of Spanish-language. Also, Tanzer et al., (1995) conducts research on native informants and asked them to deduce the accuracy of the tool and propose needed amendments. Given the recent research and notion on prejudice and procedure for test development and appraisal, it is evident that modern supporters of mean score variances on psychological tests did not followed the basic values of science. Additionally, systematic method of selecting a test instrument helps to regulate method bias. The probability of a form of bias is not on research problem but rather on distance of ethnic clusters involved. This notion defines general word for features with varied groups and pertinent to the target variable. This features statistically clarify experiential cross-cultural score variations. So, the more the characteristics display, the more likelihood of bias threats.

Most research on predictive validity demonstrates slight indication of bias and no sign of blatant prejudice on American, English-speaking minority groups. A situation where bias establishes naturally contains intercept variances that is expected of a tests less than faultless reliability. As a result, construct validity studies of the whole test produce the following deductions: (a) bias is rarely establish in particular items; (b) objects repeatedly judged by “armchair” remark (face validity) to be prejudiced demonstrate to be unprejudiced once exposed to experimental analysis (Jensen & McGurk 1987; Sandoval & Miille, 1980); (c) most assessments shows that some items display lack of prejudice than demonstrating sign of partiality; (d) experiential prejudice is repeatedly disorder in a way that content patterns or groups of examinees reprimanded is not separated; (e) experiential item bias is little and cannot elucidate the magnitude of cluster variances in the mean total scores; and (f) investigative procedures show factor uniformity across clusters and therefore, show no sign of construct bias. Cole and Moss (1989), suggested an assumption that increases the idea of assessment bias further than its statistical/procedural connotation as full mentioned in BIMT, debate. They explain that “extra-validity” is “the resolves for which a test is used, the degree to which those determinations are attained by the steps taken, and the numerous unintentional consequences and likely options to the assessment that serve similar purpose” (Cole & Moss 1989: 213).

Alternatively, research on test bias looked at changes across ethnic groups. For example, are Chinese people more reserved than British? This question needs an analysis that based on level-oriented research. Once a study employs a strategy that compares intracultural variances that transverse cultures, i.e. pre-test–post-test approach, or appraisals of sexual category across cultures, it shows that measurement unit level quality and method bias generally do not imperil the results. Only if causes of bias influence intracultural assessments differently that we can conclude that method bias impend the validity of the deductions. A typical illustration occurs when youngsters of a particular culture acquired inverse characteristics from learning, i.e. test-wiseness. On the other hand, when research aim at full-scale equality (a multicultural comparison of average scores), all prejudice is a risk to the extrapolations validity.

Messick (1989, 1996) concludes that the formation of construct, content, and criterion validity are inadequate and it must be stretched to include extra-validity worries, i.e. consequential validity. The consequential validity uses values culture and politics to examine the complete situation of a test analysis and usage (Reckase 1998). Moreover, a significant impact of BIMT that mostly ignore by test critics is the difference between prejudice and impartiality. Besides, Van de Vijver and Poortinga (1994, 1997) maintain that psychometric evaluation of bias ought to be accompanying with multicultural explanation of bias in a two added capacities: “construct and method bias.” This explains that bias in construct validity happens when “test writers across different cultures use meanings or notion that completely not overlay” (Van de Vijver & Poortinga, 1997: 30). Method bias takes place “if ethnic element not pertinent to the concept influences assessment items in a manner transversely the cultures investigated” (Van de Vijver & Poortinga 1997: 30). This happens when clusters from varied nations are examined in unrelated testing situations or are differentially acquainted with answer measures.

Conclusion and Recommendations

Evidence shows that psychological test results have the influence to change lives. Also research confirms that historical inclinations are vital for proper understanding and analysis of modern effect of psychological tests. On the other hand, research establishes that test bias did not occur in the present day. This assumption continues to increase divergent belief about the consequences of bias in psychological tests. Nevertheless, study shows that test bias is minute in assessment and it increases queries about its significance to research. This underscores the fact that information obtained in a culture is comparable to other cultures.

Research also repeatedly discloses different bias to be anticipated in a specific cross-cultural investigation and assessment method. It is revealed that that bias and its equivalence are vital to cross-cultural study and therefore variations observed in cross-cultural assessments possibly have a partly or completely different connotation than intracultural assessments. Moreover, if this emerges, then we can conclude that bias has happened in a test. A distinction about bias is made and this is subject to whether it is placed at the level of the concept (labelled construct bias), mechanism administration (method bias), or the distinct items (item bias). As frequent operators of assessment tests, psychologists always worried about the assessments of psychometric properties and this includes the visibly significant subject of bias.

A vast body of knowledge supports the corresponding validity of psychological tests, i.e. intelligent test, emotional test, personality test, to mention a few with culturally diverse groups. Therefore, psychologists ought to be self-assured in the assessments' value. Jensen (1980) concludes that "Statements of test bias and of the prejudicial use of tests cannot be overlooked by psychologists. This needs to be empirically examined through available psychometrics techniques and statistical analysis. Jensen (1980) concludes that where prejudice is established, the test should either not use on the sub-categories for which it is biased or have to be reviewed in order to remove it. Also, before the usage of tests is disallowed completely, one must reflect on another possibility to testing. This explains whether judgments established on less unbiased means of assessment (generally educational credentials, letters of reference, interviews, and biographical records) would promote fewer prejudice and larger impartiality for subgroups than outcome from the use of tests.

Recommendations

Based on various findings on bias and cultural diversity in psychological assessment, the following specific recommendations related to this paper suggest various strategies that help to reduce bias in psychological assessment.

This paper recommends that:

1. Efforts should be directed toward promoting rigorous investigation that target likely assessment bias and imprecision by using the modern and most varied methods.
2. Future research should focus on scientific expressions that encourage clearness in conversation, discussion, and problem solving.
3. Scientists and other professionals must explore and work against "armchair" face validity analysis (Jensen & McGurk 1987; Sandoval & Miille, 1980), as these impacts negatively on assessment outcome

4. In order to develop a bias free psychological test, effort should be directed towards concepts that are reliable and concur with data they planned to elucidate. This is because many defenders of the cultural testing bias hypothesis suggest assumptions and models that are not similar with existing observed data.

5. Last but not the least, significant and supportive evidence must be observed and applicable to the research study. This will prevent researchers from committing methodical blunder when administering test to diverse groups.

References

1. Anastasi, A. (1988). *Psychological testing*. New York: Macmillan.
2. Ardila, A., Rodríguez-Menéndez, G., & Rosselli, M. (Eds.). (2002). *Current issues in neuropsychological assessment with Hispanic/Latinos*. Ferraro, F. Richard. Minority and cross-cultural aspects of neuropsychological assessment. *Studies on neuropsychology, development, and cognition*. (pp. 161-179). Lisse, Netherlands: Swets & Zeitlinger Publishers.
3. Banks, S.P., Ge, G., & Baker, J., (1991). Intercultural encounters and miscommunication. In: Coupland, N., Giles, H., Wiemann, J.M. (Eds.), *Miscommunication and Problematic Talk*. Sage, Newbury Park, CA, pp. 103–120.
4. Barna, L.M., (1991). Stumbling blocks in intercultural communication. Third ed. In: Samovar, L.A., Porter, R.E. (Eds.), *Intercultural Communication: A Reader*. Wadsworth, Belmont, CA.
5. Berk, R. A. (1982), ed. *Handbook of methods for determining test bias*. Baltimore: Johns Hopkins University Press.
6. Binet, A. & Simon, T. (1973). *The development of intelligence in children*. New York: Arno Press. (Original work published 1916).
7. Boone K.B., Victor T.L., Wen J, Razani J, & Pónton M. (2007). The association between neuropsychological scores and ethnicity, language, and acculturation variables in a large patient population. *Archives of Clinical Neuropsychology*, 22(3):355–365.
8. Bochner, S., (1994). Cross-cultural differences in the self-concept. A test of Hostede's individualism/collectivism distinction. *Journal of Cross-Cultural Psychology* 25, 273–283.
9. Bracken, B. A., & McCallum, R. S. (1998). *Universal Nonverbal Intelligence Test*. Itasca, IL: Riverside.
10. Brickman A.M, Cabo R, & Manly J.J. (2006). Ethical Issues in Cross-Cultural

- Neuropsychology. *Applied Neuropsychology*, 13(2),91–100.
11. Broer, (1996). Rasch-homogene Leistungstests (3DW, WMT) im Kulturvergleich Chile–Osterreich. Erstellung einer spanischen Version einer Testbatterie und deren interkulturelle Validierung in Chile (Crosscultural comparison of the Rasch-calibrated tests 3DW and WMT between Chile–Austria and the development of a Spanish version of the test battery). University of Vienna, Austria Unpublished master’s thesis.
 12. Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since “Bias in Mental Testing.” *School Psychology Quarterly*, 14, 208–238.
 13. Cattell, J. M. (1890). *Mental tests and measurements. Mind*, 15, 373-380.
 14. Cattell, R. B. (1940). A culture free intelligence test. Part I. *Journal of Educational*.
 15. Cheung, F.M., Leung, K., Fan, R.M., Song, W.Z., Zhang, J.X., & Chang, J.P., (1996). Development of the Chinese personality assessment inventory. *Journal of Cross-Cultural Psychology* 27, 181–199.
 16. Chinese Culture Connection, (1987). Chinese values and the search for culture-free dimensions of culture. *Journal of Cross-Cultural Psychology* 18, 143–164
 17. Cohen, R., (1987). International communication: an intercultural approach. *Cooperation and Conflict* 22, 63–80.
 18. Cole, N., & Moss, P. A. (1989). Bias in test use. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 201-220). New York: Macmillan.
 19. Church, T.A., 1987. Personality research in a non-Western setting: the Philippines. *Psychological Bulletin* 102, 272–292.
 20. Cole, N. S., & Zieky, M. J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369–382.
 21. Deregowski, J.B.,& Serpell, R., (1971). Performance on a sorting task: a Ross-cultural experiment. *International Journal of Psychology* 6, 273– 281.
 22. Flanagan , D. P. Ortiz , S. O. (2001). Essentials of cross-battery assessment. NY: Wiley & Sons.
 23. Frisby, Craig L. (1998) Culture and cultural differences. In Sandoval, Jonathan et al, eds. *Test interpretation and diversity: achieving equity in assessment*. Washington, DC: American Psychological Association.
 24. Galton, F. (1869/1892/1962). *Hereditary Genius: An Inquiry into its Laws and Consequences*. Macmillan/Fontana, London.
 25. Gass, S.M., & Varone, E.M., (1991). *Miscommunication in normative speaker discourse*. In: Coupland, N., Giles, H., Wiemann, J.M. (Eds.), *Miscommunication and Problematic Talk*. Sage, Newbury Park, CA, pp. 121–145.
 26. Goodwin, R.,& Lee, I., (1994). Taboo topics among Chinese and English friends. A cross-cultural comparison. *Journal of Cross-Cultural Psychology* 25, 325–338.
 27. Gould, S. J. (1996). *The mismeasure of man* (rev. ed.). New York, NY: Norton
 28. Gregory, R. J. (2004). *Psychological testing: History, principles, and applications*. Boston: Allyn & Bacon
 29. Hammill, D. D., Pearson, N. A., & Wiederholt, J. L. (1997). *Comprehensive test of nonverbal intelligence*. Austin, TX: Pro-Ed.
 30. Hermstein, R. J., & Murray, C. (1994). The bell curve: Intelligence and class structure in American life. New York: *Free Press. sychology*, 31, 161–179.
 31. Hilliard, Asa G., III (1995), ed. *Testing African American students: special reissue of the Negro Educational Review*. Chicago: Third World Press.
 32. Ho, D.Y.F., 1996. *Filial piety and its psychological consequences*. In: Bond, M.H. (Ed.), *Handbook of Chinese Psychology*. Oxford University Press, Hong Kong, pp. 155–165.
 33. Holland, P. W., & Wainer, H. (Eds.) (1993). *Differential item functioning*. Hillsdale, NJ: Erlbaum.
 34. Holland, P.W.,& Thayer, D.T., (1988). Differential item performance and the Mantel-Haenszel procedure. In: Wainer, H., Braun, H.I. (Eds.), *Test Validity*. Erlbaum, Hillsdale, NJ, pp. 129–145.
 35. Hofstede, G., 1980. *Culture’s Consequences: International Differences in Work-related Values*. Sage Publications, Beverly Hills, CA.
 36. Hoshmand, L.T., & Ho, D.Y.F., (1995). Moral dimensions of selfhood: Chinese traditions and cultural change. *World Psychology* 1, 47–69.
 37. Hui, C.H., & Triandis, H.C., (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology* 20, 296–309.
 38. Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721–735.
 39. Irvine, S. H., & Berry, J. W. (Eds.). (1988). *Human abilities in cultural context*. New York: Cambridge University Press.
 40. Jensen, A. R., & McGurk, F. C. J. (1987). Black-white bias in “cultural” and “noncultural” test items. *Personality and Individual Differences*, 8, 295-301.

41. Jensen, A. R. (1984). Test bias: Concepts and criticisms. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 507–586). New York, NY: Plenum Press.
42. Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
43. Kagitcibasi, C., 1996. *Family and Human Development Across Cultures. A View from the Other Side*. Erlbaum, Hillsdale, NJ.
44. Klieme, E., & Stumpf, H., (1991). Dif: a computer program for the analysis of differential item performance. *Educational and Psychological Measurement* 51, 669–671.
45. Loewenstein D.A, Arguelles T, Arguelles S, & Linn-Fuentes P. (1994). Potential cultural bias in the neuropsychological assessment of the older adult. *Journal of Clinical and Experimental Neuropsychology*, 16, 623–629
46. Manly J.J. (2005). Advantages and disadvantages of separate norms for African Americans. *The Clinical Neuropsychologist*, 19,270–275.
47. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: Macmillan.
48. Messick, S. (1996). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.
49. Monnot, M. J., Quirk, S. W., Hoerger, M. & Brewer, L. (2009). Racial bias in personality assessment: Using the MMPI-2 to predict psychiatric diagnoses of African American and Caucasian chemical dependency inpatients. *Psychological Assessment*, 21 (2), 137–151.
50. Moran, M. P. (1990). The problem of cultural bias in personality assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children. Vol. 2: Personality, behavior, and context* (pp. 524–545). New York, NY: Guilford Press.
51. Mpofo, E., & Ortiz, S. O. (2009). Equitable assessment practices in diverse contexts. In E. L. Grigorenko (Ed.), *Multicultural psychoeducational assessment* (pp. 41–76). New York, NY: Springer.
52. Paranjpe, A.C., (1995). The denial and affirmation of self: the complementary legacies of east and west. *World Psychology* 1, 9–46.
53. Poortinga, Y.H., Malpass, R.S., 1986. Making inferences from cross-cultural data. In: Lonner, W.J., Berry, J.W. (Eds.), *Field Methods in Cross Cultural Psychology*. Sage, Beverly Hills, CA, pp. 17–46.
54. Reckase, M. D. (1998). Consequential validity from the test developer's perspective. *Educational Measurement: Issues and Practice*, 17(2), 13-16.
55. Reynolds, C. R. (2000a). Methods for detecting and evaluating cultural bias in neuropsychological tests. In E. Fletcher-Janzen, T. L. Strickland, & C. R. Reynolds (Eds.), *Handbook of cross-cultural neuropsychology* (pp. 249–285). New York, NY: Kluwer Academic/Plenum Press.
56. Reynolds, C. R. (2000b). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law*, 6, 144–150.
57. Reynolds, C. R., Lowe, P. A., & Saenz, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of school psychology* (3rd ed., pp. 549-595). New York: Wiley.
58. Reynolds, C. R. (1998). Race bias in testing. In R. J. Corsini & A. J. Auerbach (Eds.), *Concise encyclopedia of psychology* (2nd ed., p. 740). New York: Wiley.
59. Roid, G. H., & Miller, L. J. (1997). *Leiter International Performance Scale-Revised: Examiner's manual*. Wood Dale, IL: Stoelting
60. Sandoval, J., & Miille, M. P. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology*, 48, 249-253.
61. Sattler, J. M. (1992). *Assessment of children* (3rd ed.). San Diego: Jerome M. Sattler.
62. Scheuneman, J. (1985). Exploration of causes of bias in test items. *GRE Board Professional Report GREB No. 81–21P*, ETS Research Report 85–42. Princeton, NJ: Educational Testing Service.
63. Serpell, R., (1993). *The Significance of Schooling. Life-journeys in an African Society*. Cambridge University Press, Cambridge.
64. Serpell, R., (1979). How specific are perceptual skills? *British Journal of Psychology* 70, 365–380.
65. Shepard, L., Camelli, G., Averill, M., 1981. Comparison of six procedures for detecting test item bias using both internal and external ability criteria. *Journal of Educational Statistics* 6, 317–375.
66. Super, C.M., (1983). Cultural variation in the meaning and uses of children's "intelligence. In: Deregowski, J.B., Dziurawiec, S., Annis, R.C. (Eds.), *Explorations in Cross-cultural Psychology*. Swets, Lisse, The Netherlands, pp. 199–212.

67. Tanzer, N.K., (1995). Cross-cultural bias in Likert-type inventories: perfect matching factor structures and still biased? *European Journal of Psychological Assessment* 11, 194–201.
68. Van de Vijver, F., & Poortinga, Y. H. (1994). Bias: Where psychology and methodology meet. In A. Bouvy, F. van de Vijver, P. Boski, & P. Schmitz (Eds.), *Journeys into cross-cultural psychology* (pp. 111-126). Netherlands: Swets & Zeitlinger.
69. Van de Vijver, F & Poortinga, Y. H. (1997). Towards an integrated analysis of bias in cross-cultural assessment. *European Journal of Psychological Assessment*, 13, 29-37.
70. Van de Vijver, F. J. R., & Leung, K. (1997a). Methods and data analysis of comparative research. In J. W. Berry, Y. H. Poortinga, & J. Pandey (Eds.), *Handbook of cross-cultural psychology* (2nd ed.). (pp. 257-300). Boston, MA: Allyn & Bacon.
71. Van de Vijver, F. J. R., & Leung, K. (1997b). *Methods and data analysis for cross-cultural research*. Newbury Park, CA: Sage
72. Verney, S. P., Granholm, E., Marshall, S. P., Malcarne, V. L., & Saccuzzo, D. P. (2005). Culture-fair cognitive ability assessment: Information processing and psychophysiological approaches. *Assessment*, 12 , 303–319.
73. Yang, K.S., & Bond, M.H., (1990). Exploring implicit personality theories with indigenous or imported constructs, The Chinese case. *Journal of Personality and Social Psychology* 58, 1087–1095.

2/19/2014