

AN INVESTIGATION INTO THE RELATIONSHIP BETWEEN TEACHERS' PERCEPTIONS OF TEST BIAS AND THEIR WORKING AND TRAINING BACKGROUND

Nguyen Ngoc Nhat Minh

The University of Danang - University of Foreign Language Studies, Vietnam

Author corresponding: Nguyen Ngoc Nhat Minh - Email: nnnminh@ufl.udn.vn

Article History: Received on 17th August 2020; Revised on 02nd April 2021; Published on 17th June 2021

Abstract: This paper aims to explore the relationship between how language teachers perceive test bias and where they are working, how long they have been working, and where they were professionally trained. The data were collected from 19 in-service English teachers from Eastern and Western settings. They completed a questionnaire in which they were asked to respond to test bias stimuli and answer questions related to their teaching background and training. The stimuli contained either of two forms of bias, unfair penalization and offensiveness. Qualitative and quantitative analysis showed teachers were not fully informed of possible forms of test bias and possible ways potential biases unfairly penalize or offend students. They were better able to recognize biases of unfair penalization than offensiveness. Statistical analyses revealed teachers with over 10 years of experience were better able to recognize potential test bias than those with less experience (at 90% confidence level). The findings contribute to the current limited literature on bias in classroom language testing and assessment, leading to implications for bias review in teacher-developed assessments and teacher training.

Key words: test bias; unfair penalization; offensiveness; teachers' perceptions; training background; teaching context; work experience.

1. Introduction

Test bias has received increasing attention from educators in several last decades. As bias threatens the quality of assessments, commercial test-makers strive to eliminate bias in their testing instruments, but bias in subtler forms still persists (Wright, 2015). Despite more attention given to test bias in language classroom contexts, a careful literature review shows there is scant research both on bias in classroom assessments and on what teachers think about bias when reviewing their tests. This study helps fill a research gap by offering a better understanding of how language teachers perceive bias in testing and assessment, and thus making recommendations for teachers' test review for their own classroom.

Prior studies on how teachers perceive language assessment show that training in assessment helps teachers properly understand assessment-related issues (Mendoza & Arandia, 2009), and that teachers' work experience might be irrelevant to their perceptions of assessment (Jannati, 2015). Such results prompt a question of the relationship between teachers' perceptions of test bias and their professional training and experience. To explore the issues mentioned above, the current study surveyed in-service English teachers working in Eastern and Western environment to address the following research questions:

- 1) How do teachers perceive bias in testing and assessment? and
- 2) How do teachers' training context, teaching context, and work experience influence their perceptions of bias?

Cite this article as: Nguyen, N. N. M. (2021). An investigation into the relationship between teachers' perceptions of test bias and their working and training background. *UED Journal of Social Sciences, Humanities and Education*, 11(1), 1-11.

2. Literature Review

Test bias refers to systematic error penalizing a test-taker group's performance (Shephard, Camilli, & Averil, 1981). Such bias emerges from internal components of an assessment procedure (Elder, 2012), such as invalid test elements, which systematically leads to better and worse results for certain groups of examinees (Standards for Educational and Psychological Testing, 1999, as cited in Educational Testing Service's Smarter balanced assessment consortium: Bias and Sensitivity Guidelines, 2012).

One useful approach to understanding assessment bias is that bias can come in two forms: *offensiveness* and *unfair penalization* (Popham, 2014). The former results from stereotypes of a group of test takers for whom the assessment is intended, which may cause them to take offense, getting distracted from the question and fail to perform well. The latter stems from an element disadvantaging one student group over others, because that group lacks the information necessary to answer the question. This distinction of the two forms shaped the operationalized construct of test bias in the current study.

In regard to sources of assessment bias, researchers state bias might offend or penalize some groups of test takers based on their ethnicity, gender, socioeconomic status (Elder, 2013), first language background, and their background knowledge (Reynolds & Suzuki, 2003). The Bias and sensitivity guidelines (Educational Testing Service, 2012) provides a specific list of topics that might cause people to feel offended and upset (for example, abuse of people or animals, killings of animals for sport, sexual behavior, torture, stereotypes, group labels, and so on) and those that have become increasingly controversial (for instance, euthanasia, gun control, or ethnic conflicts), which should be avoided in designing assessments. The guidelines also suggest some topics that demand caution in stimulus design, such as antisocial or inappropriate behaviors (bullying, cheating, etc.), family problems, or descriptions of luxury, and many others.

Most of the studies conducted on assessment bias have examined the issue in large-scale, standardized tests, and only in the interpretation and evaluation of test-takers' scores (Educational Testing Service, 2012;

Elder, 2012; Mellenbergh, 1989; Wright, 2015). Few published works consider bias in small-scale classroom assessment. Those few sources mostly suggest techniques for teachers to avoid biases in informal classroom assessments. These techniques include carefully selecting assessment methods and materials; avoiding assignments that are potentially biased towards students who have higher academic investment; using scoring rubrics; and consulting with a colleague who comes from the same subgroup as that of the students before a test is used (Blankenship, Hubbard, & Johnson, 2009; Popham, 2014).

There are also few studies examining language teachers' perceptions of classroom assessment which are mostly concerned with what teachers think about types of assessment and about their uses in the classroom (Jannati, 2015; Mendoza & Arandia, 2009; Shim, 2009; Sikka, Nath, & Cohen, 2007). None specifically mention bias, except briefly discussing teachers' criteria to maintain clear assessment procedures. However, understanding teachers' perceptions of test bias and raising their awareness of potential biases in teacher-made assessments are important to our effort to create tests with absence-of-bias (Popham, 2014) because perceptions and attitudes greatly affect their assessment practices (Davison, 2004).

Reporting that test bias has been less prevalent than it was some decades ago due to measurement specialists' attempts to mitigate bias, Popham (2014) nonetheless points out "systematic attention to bias eradication" is not common in "teacher-developed assessment procedures" (p. 146). He adds that if teachers, as test developers, can realize forms of potential assessment bias which affects learner performances, they can monitor and reduce bias in their own classroom tests. Exploring teachers' ability to monitor bias in this way is precisely the aim of the current study.

3. Methodology

3.1. Participants

The participants were nineteen English language teachers currently teaching English (either in ESL or EFL contexts) and having developed tests and assessments for their students by the time they answered

the research questionnaire. Most of them received training in language assessment with little coverage of test bias. Among those reporting where they learned those skills and knowledge, eight were trained in western contexts (the US, England, Australia), and four in eastern contexts (Vietnam and China).

3.2. Materials

The questionnaire was created online with Google Forms, then piloted with two teachers who were members of the target research population before reaching potential participants via email on Google Form and a TESOL Facebook page. Nineteen teachers submitted complete responses within two weeks.

To clarify how teachers understand test bias, the questionnaire was designed with two parts. The first part of the survey explored background information about the teachers, including their experience, working and training context. The second part examined participants' perceptions of test bias by exposing them to testing situations and test items (which were referred to as *stimuli* in this research). These stimuli (see *Appendix*) were concerned with two presented forms of assessment bias, *unfair penalization* and *offensiveness*.

There were more stimuli containing bias of offensiveness than of unfair penalization. A potential bias of the latter form might unfairly penalize test-takers

in terms of the assessed content to which they were not exposed, which is often the case. An offensive test bias, however, might offend people in different ways. Therefore, the designed offensive biases fall into certain categories adapted from the Educational Testing Service's Smarter balanced assessment consortium: Bias and Sensitivity Guidelines (2012). Although there are many other categories, seven were chosen for offensive bias stimuli in this study, namely, children abuse, description of luxury, stereotype, (offensive) group labelling, gender, gruesome detail, and animal abuse.

The stimuli included four test situations and six test items taken directly or paraphrased from the Educational Testing Service's Guidelines for Fairness Review of Assessments (2009), the Educational Testing Service's Smarter balanced assessment consortium: Bias and Sensitivity Guidelines (2012), and Popham's (2014) book "Classroom assessment: What teachers need to know." Each stimulus represented an *intended bias* (intentionally selected). In several stimuli, the intended biases were considered *acceptable* or *unacceptable*, indicating they were reasonably or unreasonably biased in some circumstances. Each intended bias comprises two aspects measuring participants' perceptions, i.e., ability to recognize the bias and ability to explain the bias. These terms are illustrated as follows (for full explanation, see *Appendix*):

Table 1. Examples of Intended Bias

Stimulus	Intended bias	
	Form of bias	Content
[testing situation] A reading test contains complex language structures and is intended to measure students' ability to read challenging materials.	Unfair penalization	The potential bias is considered <i>acceptable</i> : Although the complex structures used in the test might disadvantage certain student groups, it can be considered not biased or the potential bias, if any, is not severe at all because the difficulty level correlates with the test purpose.

3.3. Analysis

Concerning how the teachers perceived the stimuli, each stimulus was given a name code. For example, the example mentioned above was coded as "complex structures" (see *Appendix*). The raw data were color-

coded to see if the teachers' perceptions (which can be specified as teachers' ability to recognize and to explain the bias) matched the intended biases and were considered *correct* responses if they did. Next, the data were organized into tables like the one as follows:

Table 2. Example of Correct/Incorrect Participants’ Perceptions to a Stimulus

Stimulus code name	Response	Interpretation	Correct?
Complex structures	“This does not really have any bias. It clearly shows the objective of the test.”	This participant saw the bias but understood the test purpose should be prioritized although some tested knowledge is unknown to some test-takers, which helps ensure test validity.	Yes
	“There are many ways to measure students’ ability to read challenging materials... it may not be fair to include complex language structures...”	This participant did not recognize the bias, saying complex structures should not be used.	No

The number of correct/incorrect responses are the main data source to verify whether the participants were able to perceive potential biases. The higher percentage shown in their answers to the stimuli, the more able they were considered to recognize and/or explain potential biases. The percentages of correct responses also indicate which intended biases teachers might be more able to recognize and become aware of.

To explore the relationship between teachers’ three background factors (independent variable), which include training context (eastern versus western), teaching context (ESL versus EFL) and work experience, and their perceptions towards the intended biases, two inferential statistics - Wilcoxon Rank Sum and Kruskal Wallis - were used. The statistics could show any significant difference in the number of correct responses (dependent variable) given by the participants if they gain language assessment training in an eastern or western context, if they are teaching in an ESL or EFL setting, and if their experience is limited (< 4

years), moderate (4 – 10 years), or rich (> 10 years). Alpha was set as 0.1 (90% certainty) for all analyses to not miss any possible differences between the variables. If the results are smaller than 0.1, there is a significant difference between different groups of teachers.

4. Results and Discussion

4.1. Teachers’ perceptions towards test bias

Towards bias of unfair penalization

Table 3 shows within each stimulus, the participants were more able to identify the intended biases of unfair penalization than able to explain them. Regarding the stimulus about technical knowledge and the one about complex structure, more than half of the teachers recognized the intended biases (78.95% for the former and 63.16% for the latter), but fewer could explain the biases seen (47.37% and 26.31%, respectively). The stimulus “opera” witnessed an equal percentage of teachers (36.84%) in both being able to identify and explain the bias.

Table 3. Percentage of Participants Giving Correct Responses to Unfair Penalization Intended Biases

	Stimulus		
	Complex structure	Technical knowledge	Cultural knowledge/Opera
	%	%	%
Ability to recognize bias (N = 19; mean = 11.3)	63.16 (n = 12)	78.95 (n = 15)	36.84 (n = 7)
Ability to explain bias (N = 19; mean = 7)	26.31 (n = 5)	47.37 (n = 9)	36.84 (n = 7)

Across all three stimuli, the teachers were most able to recognize the bias concerning technical knowledge (78.95%) and least able in cultural knowledge (36.84%). They were also most able to explain the issue of technical knowledge (47.37%), but least able in that of complex structure (26.31%). Generally, the technical knowledge stimulus showed the highest percentages for the teachers' ability to perceive the intended bias as they assumed technical content might be too specialized for students to understand.

The stimuli about complex structures and cultural knowledge were less visible to the teachers since whether these items are potentially biased actually depends on contexts in which the test is used. Specifically, the complex structure stimulus which is potentially biased is designed to match the test purpose, i.e., measuring learner ability to read challenging materials. The "opera" stimulus might be biased or not

depending on whether the cultural knowledge of opera was provided.

Towards bias of offensiveness

According to Table 4, almost all stimuli of offensive bias witnessed equal proportions between teachers' ability to recognize the intended bias and their ability to explain it. For example, around 42% of the teachers could both see the bias on spanking children and explain how the stimulus might be biased. Only the "wheelchair" stimulus showed somewhat difference between the participants' ability to identify and ability to explain the bias (15.79% versus 26.32%, respectively). This difference probably means the respondents saw the negative connotation in portraying people with disabilities but were not bothered about such negative group labelling, and therefore did not see how the item could be biased.

Table 4. Percentage of Participants Giving Correct Responses to Offensive Intended Biases

Categories of offensiveness	Children abuse	Description of luxury	Stereotype	Group labelling	Gender	Gruesome detail	Animal abuse
Stimulus	Spanking children	Expensive tech	French arrogance	Wheelchair/ "restricted"	Girl winning	Slavery	Animal experiment
	%	%	%	%	%	%	%
Ability to recognize bias (N = 19; mean = 5.29)	42.11 (n = 8)	5.26 (n = 1)	31.58 (n = 6)	15.79 (n = 3)	57.89 (n = 11)	21.05 (n = 4)	21.05 (n = 4)
Ability to explain bias (N = 19; mean = 5.57)	42.11 (n = 8)	5.26 (n = 1)	31.58 (n = 6)	26.32 (n = 5)	57.89 (n = 11)	21.05 (n = 4)	21.05 (n = 4)

The results also show the highest percentage of teachers recognizing the intended bias (57.89%) for the "girl winning" stimulus and much lower percentages for the stimuli of "wheelchair" (21.05%), "slavery" (21.05%), "animal experiment" (15.79%), and especially "expensive tech" (5.26%). A similar pattern also repeated for teachers' ability to explain the biases (with approximate percentages): Teachers were most able to explain the bias "girl winning" and much less

able concerning those of "slavery," "animal experiment," and "expensive tech."

Generally, the respondents were most likely to recognize offensive bias regarding *gender*, but least likely related to *group labelling*, *gruesome detail*, *animal abuse*, and *description of luxury*. The pattern can be interpreted by considering the participants' social and cultural background. Probably in some cultures where the teachers come from, such topics of group

labelling, gruesome detail, animal abuse, and description of luxury are simply factual incidents, and therefore these teachers did not feel bothered about them. Gender bias, compared to other categories of offensiveness, might be most visible and a familiar bias type to teachers as gender equality is discussed very often these days.

4.2. The Relationship Between Teachers’ Background Factors and Their Perceptions

Both descriptive and inferential statistical analyses (see Tables 5, 6, and 7) were conducted to examine relationships between the teachers’ background factors and their perceptions of bias (based on the participants’ correct responses to the bias stimuli).

- There is no statistically significant difference between the number of the teachers’ correct responses and the context (eastern or western) where they received training in language assessment ($W = 8, p = .14, \alpha = .1$). There is no relationship between the teachers’ training context and their ability to perceive potential test bias.

Table 5. Descriptive Statistics of two Group Trained in an Eastern or Western Context

Descriptive statistics	min	max	median	mean	sd
east.trained (n = 4)	0	12	3.5	4.75	5.25
west.trained (n = 9)	0	16	11	10.22	4.68

- There is no statistically significant difference between the number of the teachers’ correct responses and the context (ESL or EFL) where they are teaching ($W = 35.5, p = .22, \alpha = .1$). There is no relationship between the teachers’ working context and their ability to perceive potential test bias.

Table 6. Descriptive Statistics of two Group Working in an ESL or EFL Context:

Descriptive statistics	min	max	median	mean	sd
esl.teach (n = 3)	5	14	12	10.33	4.73
epl.teach (n = 16)	0	16	6.5	6.438	5.37

- Although Kruskal-Wallis analysis did not show any significant differences between the three groups of people who have limited, moderate, and rich work experience (Kruskal-Wallis chi-squared = 2.95, $df = 2, p = 0.23, \alpha = .1$), post-hoc analyses with Wilcoxon Rank Sum reported a 90% certainty of a significant difference between the number of the teachers’ correct responses and work experience of more than 10 years ($W = 11, p = .07, \alpha = .1$). It was concluded that teachers with more than 10 years of experience might be more able to recognize potential test bias than the other two groups, which can be observed clearly in Figure 1.

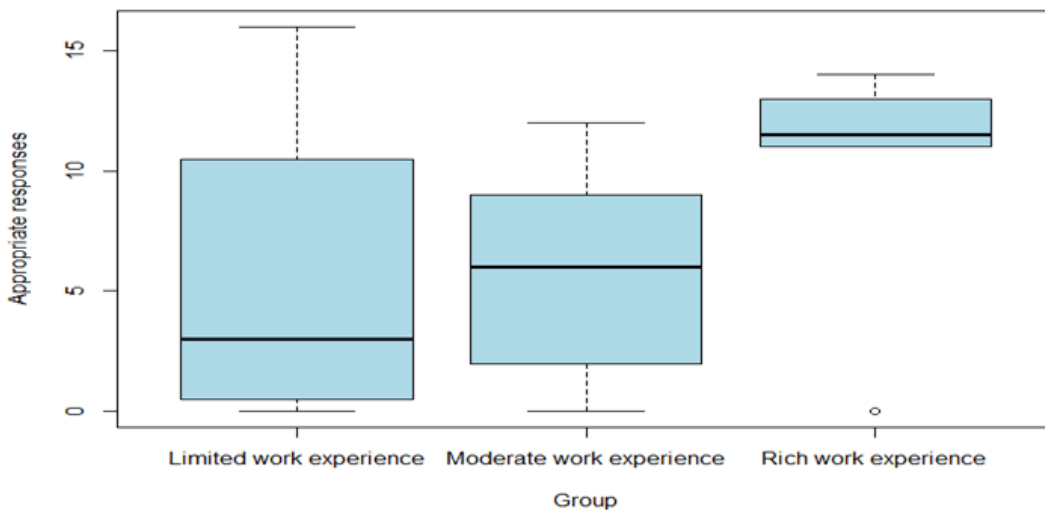


Figure 1. Plot of the three groups and their numbers of correct responses

Table 7. Descriptive Statistics of the Groups Whose Work Experience are Limited, Moderate and Rich

Descriptive statistics	min	max	median	mean	sd
limited.experience n = 4	0	16	3	5.5	7.33
mod.experience n = 9	0	12	6	5.67	4.15
rich.experience n = 6	0	14	11.5	10.17	5.12

5. Conclusion

The current study began with the interest in how language teachers perceive bias in testing and assessment, and the relationship between their perceptions with their professional background (training context, teaching environment, and work experience). The findings indicated the teachers were not fully informed of possible forms in which test bias could come and possible ways potential biases might have penalized or offended students.

In order to improve their own understanding of test bias, teachers first need to be aware that bias should be reviewed after test items have been written and before the test is officially used. They also need being informed of the two forms from which test bias might come, which are, *unfair penalization* and *offensiveness*. In addition, teachers should be aware of ways helping mitigate biases. Those techniques include using scoring criteria, sharing with students the test format and possible topics and skills, having tests peer-reviewed, planning test accommodations, and seeking student feedback to test design. Informed of such aspects of the “absence-of-bias” and bias mitigation, teachers can “routinely ... judge their own assessments and those educational assessments developed by others” (Popham, 2014, p. 146). The result discussion revealed substantial work experience (more than 10 years) might enable teachers to be more able to detect potential test bias, which suggests substantial practice (including teaching and assessment practices) can help. Therefore, regular practice in reviewing their own assessments for bias can empower teachers in increasing the quality of their classroom assessment.

The pedagogical implications discussed above can be more impactful if shared with other professionals, who look forward to improving their test review, through teacher training workshops or seminars with the topic of test bias and bias mitigation included in the Language Assessment component of teacher education programs, and professional conferences in the field. The implications, although drawn from the research on English language teachers, are applicable to teachers developing and reviewing tests of other subjects delivered in the English language (e.g. math tests in English) and even tests of other languages.

Although conducted with a relatively small sample size, this study has made certain contributions to the current limited research on test bias within classroom setting. As mentioned in the literature review, there is little research on bias in classroom assessment, on teachers’ perceptions towards the issue, as well as on the relationship between their perceptions and their training or experience. Therefore, the findings presented above have supported the current literature in those aspects.

References

- Blankenship, J., Hubbard, B., & Johnson, D. (2009, April 27). *Bias in classroom assessment*. Retrieved from <https://goo.gl/gldL0T>
- Davison, C. (2004). The contradictory culture of teacher-based assessment: ESL teacher assessment practices in Australian and Hong Kong secondary schools. *Language Testing*, 21(3), 305-334. <https://doi.org/10.1191/0265532204lt286oa>
- Educational Testing Service. (2009). *ETS guidelines for ets review of assessments*. Retrieved from https://www.ets.org/Media/About_ETS/pdf/overview.pdf
- Educational Testing Service. (2012). *Smarter balanced assessment consortium: Bias and sensitivity guidelines*. Retrieved from <https://portal.smarterbalanced.org/library/en/bias-and-sensitivity-guidelines.pdf>
- Elder, C. (2012). Bias in language assessment. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 406–412). Oxford: Wiley-Blackwell.

<https://doi.org/10.1002/9781405198431.wbeal1198>

- Jannati, S. (2015). ELT teachers' language assessment literacy: Perceptions and practices. *The International Journal of Research in Teacher Education*, 6(2), 26-37.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, 13(2), 127-143. [https://doi.org/10.1016/0883-0355\(89\)90002-5](https://doi.org/10.1016/0883-0355(89)90002-5)
- Mendoza, A.A.L., & Arandia, R.B. (2009). Language testing in Colombia: A call for more teacher education and teacher training in language assessment. *Profile*, 11(2), 55-70. Retrieved from <http://revistas.unal.edu.co/index.php/profile/article/view/11442/36795>
- Popham, W. J. (2014). *Classroom assessment: What teachers need to know* (7th ed.). Boston, MA: Pearson.
- Reynolds, C. R., & Suzuki, L. A. (2003). Bias in psychological assessment: An empirical review and recommendations. In I. B. Weiner, J. R. Graham, & J. A. Naglieri (Eds.), *Handbook of Psychology* (Vol. 10, pp. 82-113). Wiley Online Library. <https://doi.org/10.1002/0471264385.wei1004>
- Shepard, L., Camilli, G., & Averil, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational and Behavioral Statistics*, 6(4), 317-75. <https://doi.org/10.3102/10769986006004317>
- Shim, K. N. (2009). *An investigation into teachers' perceptions of classroom-based assessment of English as a foreign language in Korean primary education*. (Doctoral dissertation). Retrieved from the database of University of Exeter.
- Sikka, A., Nath, J., & Cohen, M. (2007). Practicing teachers' beliefs and uses of assessment. *International Journal of Case Method Research & Application*, 10(3), 239-253.
- Wright, W. E. (2015). *Foundations for teaching English language learners: Research, theory, policy, and practice* (2nd ed.). Philadelphia, PA: Caslon.

NGHIÊN CỨU MỐI LIÊN QUAN GIỮA NHÌN NHẬN CỦA GIÁO VIÊN VỀ THIÊN VỊ TRONG ĐÁNH GIÁ VỚI NỀN TẢNG CHUYÊN MÔN VÀ KINH NGHIỆM CỦA HỌ

Nguyễn Ngọc Nhật Minh

Trường Đại học Ngoại Ngữ - Đại học Đà Nẵng, Việt Nam

Tác giả liên hệ: Nguyễn Ngọc Nhật Minh - Email: nnnminh@ufl.udn.vn

Ngày nhận bài: 17-8-2020; ngày nhận bài sửa: 02-4-2021; ngày duyệt đăng: 17-6-2021

Tóm tắt: Bài báo này nhằm trình bày mối liên quan giữa cách nhìn nhận thiên vị trong đánh giá của giáo viên và các yếu tố về nền tảng chuyên môn và kinh nghiệm của họ. Dữ liệu nghiên cứu được thu từ 19 giáo viên đã và đang giảng dạy tiếng Anh ở môi trường phương Đông và phương Tây. Các giáo viên này tham gia trả lời một khảo sát về các thông tin liên quan đến chuyên môn và kinh nghiệm giảng dạy của họ, và được yêu cầu phản hồi lại các câu hỏi có chứa thành tố thiên vị trong đánh giá. Các câu hỏi này chứa các dạng thiên vị thuộc hai thể loại: bài trừ, và xúc phạm. Dữ liệu nghiên cứu được phân tích định tính và định lượng. Kết quả cho thấy các giáo viên chưa hiểu hết về các loại hình thiên vị, và cũng chưa nhìn nhận được hết những tác động của thiên vị lên người học trong kiểm tra đánh giá. Tuy vậy, các giáo viên nhận ra loại hình thiên vị "bài trừ" tốt hơn loại hình thiên vị "xúc phạm". Các bước phân tích định lượng cho thấy giáo viên với hơn 10 năm kinh nghiệm làm việc có khả năng nhận ra thiên vị trong đánh giá tốt hơn (ở mức 90% chắc chắn). Kết quả nghiên cứu có ý nghĩa đóng góp vào tổng quan lý thuyết còn hạn chế hiện tại về kiểm tra đánh giá trong lớp học, và có những đề xuất cho quy trình kiểm duyệt đề thi của giáo viên và cho việc tập huấn giáo viên trong lĩnh vực này.

Từ khóa: thiên vị trong đánh giá; thiên vị "bài trừ"; thiên vị "xúc phạm"; nhận thức của giáo viên; đào tạo chuyên môn; môi trường dạy; kinh nghiệm.

Appendix

Intended use of the survey testing situations and test items

<i>Test component</i>	<i>Stimulus name codes</i>	<i>Form of potential bias</i>	<i>Intended use</i>
<i>Testing situations</i>			
An argumentative writing topic asks students to determine the degree to which they agree with a controversial issue, like spanking children as a way to teach them about right and wrong.	Spanking children	Offensiveness	The potential bias is considered <i>unacceptable</i> because: Spanking children is an unacceptable ethical issue in some societies or cultures, so forcing the supporting view of this issue on test takers might upset members coming from those cultures. In other words, the item might offend certain groups of test takers and contain bias in the form of offensiveness, accordingly.
A reading test contains complex language structures and is intended to measure students' ability to read challenging materials.	Complex structure	Unfair penalization	The potential bias is considered <i>acceptable</i> because: Although the high complexity of the structures used in the reading test might disadvantage certain groups of students, it can be considered not biased or the potential bias, if any, is not severe at all because the level of difficulty correlates with the test purpose.
A writing task requires some knowledge about how tools and machines work or are assembled in a process of how a product is made.	Technical knowledge	Unfair penalization	The potential bias is considered <i>unacceptable</i> because: The test requires students to have some specialization knowledge of how certain machines work to accomplish the task. In this situation, the test might unfairly penalize certain subgroup of test takers.
An assessment procedure requires students to first read a text about how students in a private middle school use expensive, cutting-edge technologies in their daily life and then to write a short paragraph about the benefits of owning such a technology.	Expensive tech	Offensiveness	The potential bias is considered <i>unacceptable</i> because: Although the students can gain some background knowledge of how to use expensive technologies in daily life to start their writing, descriptions of luxuries like "private school," "expensive, cutting-edge ...in daily life" can upset students with limited financial conditions or low-quality living standards.

<i>Written test items</i>				
The character delivering the monologue attributes the arrogance of French people to which of the following?	French arrogance	Offensiveness		The potential bias is considered <i>unacceptable</i> because: Describing all of the people in a nation as “arrogant” is a clear case of offensive stereotyping. In other words, the item might offend certain groups of test takers and contain bias in the form of offensiveness, accordingly.
In the play, Luz was restricted to a wheelchair for what reason?	Wheelchair	Offensiveness		The potential bias is considered <i>unacceptable</i> because: The phrase “was restricted to a wheelchair” should be replaced with more objective terminology such as “began using a wheelchair.” In other words, the item might offend certain groups of test takers (who support or have membership with people with physical disabilities) and contain bias in the form of offensiveness, accordingly.
Surprisingly, a girl won the math contest.	Girl winning	Offensiveness		The potential bias is considered <i>unacceptable</i> because: By expressing surprise that a girl won the math contest, the excerpt reinforces the stereotype that girls have less quantitative ability than boys. In other words, the item might offend certain groups of test takers and contain bias in the form of offensiveness, accordingly.
Wagner used the orchestra to achieve certain effects in much the same way that other composers of operas used the singers.	Opera	Unfair penalization		The item might contain potential bias, depending on the specific content provided in the reading passage: <ul style="list-style-type: none"> - Bias might exist if the knowledge needed to answer the questions was included in the passage. - Bias might not exist (or the potential bias, if any, is slightly severe) if understanding the passage required knowledge of opera and how composers “used” the orchestra or singers. The potential bias might therefore unfairly penalize some test takers depending on the context.
Read the excerpt from the diary of a ship captain engaged in transporting slaves and watch the	Slavery	Offensiveness		The potential bias is considered <i>unacceptable</i> because: Mention of slavery as a topic is acceptable but forcing test takers to imagine that they personally experienced the

video dealing with the history of slavery in the United States. Imagine that you are a newly captured slave. Describe your experiences on land and on the sea during your journey from Africa to the United States. Use information from both the diary and the video in your description.

transatlantic journey, during which many captives are known to have suffered and died, will be upsetting to some students. In other words, the item might offend some groups of test takers and contain bias in the form of offensiveness, accordingly.

Harlow was best known for the experiment in which he separated infant monkeys from their mothers shortly after the infants were born.

Animal
experiment

Offensiveness

The potential bias is considered *unacceptable* (in language and arts assessment) because: The excerpt includes painful experimentation, which might upset people who are for animal rights. This, however, would be acceptable in a psychology test.
