The Journal of Asia TEFL

Fairness and Justice in L2 Classroom Assessment: Perceptions from Test Takers

Matthew P. Wallace

University of Macau

Fairness and justice are key aspects in the evaluation of language assessments. In this paper, I argue that fairness and justice are subjective perceptions of test takers. Consistent with Kunnan (2018), fairness was conceived as a test quality, while justice was a quality of the social entity administering the test; which in this study was the language program. Test fairness was conceptualized as a multi-dimensional construct—*distributive* fairness (how fairly learners perceive their score relative to their effort) *procedural* fairness (equality of procedures administered during the assessment), *interactional* fairness (respectful interpersonal treatment of the students by the test administrator), and *informational* fairness (how fairly information about the assessment and its procedures were provided). In total, 83 university L2 learners in Taiwan completed an online questionnaire gauging perceptions of a single test administration. Regression analysis revealed that *procedural* and *interactional* fairness were predictive of justice perceptions of the language program. *Distributive* fairness shared a relationship with justice perceptions but was not predictive. These findings suggest that participants made justice judgments of their language programs based on how fairly the procedures used to administer a single test were carried out and how respectfully the instructor administering the test treated them.

Keywords: fairness, justice, L2 classroom assessment

Introduction

Fairness and justice have for some time now been considered key aspects in the evaluation of language assessments. When developing a test, validating its scores and use, and revising it for subsequent administrations, concerns of fairness and justice are paramount. Of the two constructs, fairness has received most of the attention in the recent past, though justice has received increased interest of late (e.g., Kunnan, 2018; McNamara & Ryan, 2011; Song, 2016). Currently, language tests are considered to be fair or just based on the results of an objective analysis of the test, its administration, and subsequent use of test scores from the test (McNamara & Ryan, 2011). Specifically, Kunnan (2018) argues that if the analysis reveals an absence of bias on the test, sufficient evidence of construct validity, and ethical use of scores, then the test is considered fair and the organization administering the test to be just. This approach implicitly assumes that fairness and justice are objective qualities, which are determined by those who are not directly involved in the testing process, nor are they affected by the administration or use of the test scores by the testing institution. This is an important omission, since the potential effects of unfair testing practice can have negative consequences for test takers and possibly testing institutions that administered the test (Chory, 2007). The perspective taken in this study is that test fairness and institutional justice are subjective in nature, where a test and institution are considered fair and just, respectively, if the test takers

themselves perceive them to be. This addresses a need in the literature called for by Kunnan, that perspectives of stakeholders of the test be taken into account.

When investigating fairness and justice in language testing, research has primarily focused on the administration of standardized tests (e.g., Song, 2016). While it is certainly important that standardized tests be carried out in a fair way due to the high stakes nature of the assessments, it is equally important for classroom assessments to be carried out fairly, since most language assessments are done in the classroom. Though classroom assessments are considered to be of lower stakes, they too have important consequences if not administered fairly because they may also dictate grades received for a course. For example, admission to university or other academic institutions may depend not only upon performance on standardized language tests (i.e., TOEFL or IELTS), but also performance in the classroom. If a student perceives his language assessment to be unfair, he may lose motivation to learn, which in turn may harm his language acquisition and ultimately, academic performance. To gain insight into these learner perceptions, the current study aimed to investigate fairness and justice in the classroom setting. Before reviewing the classroom justice literature, the fairness and justice literature in language testing will be discussed.

Fairness and Justice in Language Testing

Fairness has been frequently discussed with regard to its relationship with validity argumentation (e.g., Karami, 2013; McNamara & Ryan, 2011; Xi, 2010). In describing the relationship, Xi (2010) identified three approaches. First, fairness and validity can be conceived as separate qualities of a test and the relationship between the constructs then is ambiguous. Second, fairness can be conceived as an overarching test quality that subsumes validity. From this approach, evidence of validity is used to support claims that a test administration and its outcome(s) were fairly used. The final approach conceives fairness and validity as being directly linked and arguments of validity require evidence of fairness to be included. Absent from much of this discussion is the construct of justice, but this is likely due to concepts of fairness and justice often used synonymously in the literature. Noticing this trend, McNamara and Ryan (2011) attempted to distinguish between the constructs within the context of validity. They proposed that fairness and justice occupy different aspects of Messick's (1989) validity matrix. According to them, fairness is associated with the *Evidential Basis* of test score interpretation and use, where construct validity, procedural equality, and the technical qualities of the test (psychometric properties) are key sources of validity evidence. In contrast, justice is associated with the *Consequential Basis* of score interpretation and use, where the values implicit within the test constructs and the social consequences of using the scores serve as sources of validity.

Another distinction between the two constructs have been by Kunnan (2000, 2004, 2018). Refining his definitions of fairness and justice from earlier conceptions of the construct (e.g., Kunnan, 2004), Kunnan (2018) lists two principles of language assessment fairness and justice based on insights from moral philosophy: the principle of fairness and the principle of justice. The principle of fairness has the following sub-principles: (1) it provides learners with adequate opportunities to learn, (2) the score interpretations are consistent and meaningful for all test takers, (3) bias and construct irrelevance are avoided, and (4) there is equitable treatment of test takers by ensuring adequate access, administration, and standard setting procedures. The principle of justice, in contrast, has the following sub-principles: (1) a test ought to foster beneficial consequences in society, and (2) promote positive values through public reasoning of the assessment. Kunnan distinguishes between the two principles broadly by stating that fairness is related to test takers, whereas justice is related to institutions that design, develop, administer, and use a test.

The current study takes a similar position regarding fairness and justice, in that fairness refers to perceptions of test takers during a testing event and justice refers to perceptions of the social entity administering the test. The social entity examined in this study is the language program responsible for delivering the assessment. Drawing from the classroom justice literature (described below), a testing

event is considered fair if learners view (1) the procedures that are used throughout the event to be equal for all test takers, (2) the score to fairly represents the learners' performance, (3) the communications with the language instructor to be respectful, and (4) the information about the test and score to be timely, specific, and accurate. The language program administering the test is considered just if learners perceive it to be fair.

Because fairness and justice are conceptualized as being subjective perceptions, the role of culture on these perceptions may play an important role. Though research on this is scant, a recent study conducted in a Chinese university context may provide some insight into how test takers in this context may view test fairness. Song (2016) reported that language learners and language instructors felt scores from a standardized language test were acceptable (1) if the procedures were administered equally for all test takers, (2) if the outcomes of test performance were based on merit, (3) if the test adequately discriminated among test takers and feedback was tailored to the test taker, (4) if test takers had adequate familiarity with the testing-oriented culture, and (5) if the test was administered in a practical manner so as not to waste resources. Though participants in Song's study reflected on a standardized test administration, the findings provide useful insight into how learners in the current study may have viewed the fairness of their classroom test. The first three aspects were examined in the current study, so the results may provide corroborating evidence to support these earlier findings. The current study moved beyond perceptions of the testing event itself to include perceptions of the social entity administering the test. The results of this investigation should provide useful feedback for language instructors and language programs, alike, as they may show how students perceive their test administration.

Despite an increase in the amount of attention fairness and justice in language assessment have received of late, what remains unclear is how fairness of a test is related to justice of the social entity administering the assessment. To address this gap, the current study drew upon the classroom justice literature to examine the relationship between the constructs. Drawing on this body of literature also addresses the need for evidence of subjective perceptions of assessment fairness and justice from the perspective of the language learner in a classroom environment.

Dimensions of Classroom Assessment Fairness

Classroom justice refers to the fairness perceptions of educational assessment outcomes and the procedures that are employed to arrive at those outcomes (Chory-Assad & Paulsel, 2004). Emerging from the organizational justice literature (e.g., Byrne & Cropanzano, 2001; Colquitt, 2001), classroom justice has been conceptualized to include four dimensions: *distributive* justice, *procedural* justice, *interactional* justice, and *informational* justice. To remain in line with the language assessment literature, these dimensions will be described as dimensions of fairness, as opposed to justice, as they refer to appraisal of a single testing event. *Distributive* fairness refers to perceptions of a given outcome being fair (Deutsch, 1985). Underlying this aspect of fairness is equity theory (Adams, 1965), which posits that people perceive a relationship to be fair when the rewards they earn from a given transaction are equitable to the contributions they make (Hatfield, Rapson, & Bensman, 2012). In language assessment, *distributive* fairness is fostered when students perceive their test score as being a fair representation of their performance.

Procedural fairness refers to fairness of the procedures used to arrive at an outcome (Byrne & Cropanzano, 2001; Leventhal, 1980). This aspect of fairness emerged from research highlighting the importance of the processes through which resources are distributed in an organization (Lind & Tyler, 1988; Thibaut & Walker, 1975). In classrooms, *procedural* fairness involves the fairness of the methods used to assign grades or award scores on a test. For example, an instructor demonstrates fair procedures if all students in a class are evaluated using the same scoring criteria that is applied consistently for every learner. Utilizing separate scoring criteria for certain learners or inconsistently applying the criteria would be considered procedurally unfair practice. This aspect of justice has received the most attention in the language testing literature, though the *procedural* fairness term has not been used to describe the

phenomenon. Kunnan (2018), McNamara and Ryan (2011), Song (2016), and Xi (2010) have all noted that fairness in language assessment involves ensuring the procedures in which an assessment is administered are applied equally for all test takers. This includes using statistical analysis to verify absence of bias in the test items and validate that the test measured its intended construct(s). In language testing, *procedural* fairness is fostered when students perceive the procedures in which a test is administered as being equally applied for all test takers.

Interactional fairness refers to fairness of the interpersonal treatment of students by teachers in the classroom context (Chory, Horan, & Houser, 2017). This aspect of classroom fairness is based on the conception that interpersonal relations influence the distribution of resources (Bies & Moag, 1986). In the classroom, this refers to the grades awarded by the instructor. In language testing, *interactional* fairness is fostered when the social entity administering a test (e.g., instructor, language program, testing organization, etc.) communicates with the test taker in a fair, respectful manner throughout a testing event; including prior to, during, and after test administration. This aspect of fairness has yet to be adequately addressed in the language testing fairness literature, though it has been identified by teachers as an important aspect of test fairness (Tierney, 2014). The current study aims to overcome this deficiency by including *interactional* fairness as a key construct influencing justice judgments of a language program.

Informational fairness refers to fairness of the information of the procedures and outcomes of an evaluative event. This dimension has yet to be included with the other three aspects of classroom fairness, though it has been identified as an important dimension of organizational justice (e.g., Colquitt, 2001). Originally part of *interactional* fairness, *informational* fairness is viewed as a separate dimension (Greenberg, 1993). It is based on part of the Bies and Moag (1986) conception of *interactional* justice that fair communications about an event are reasonable, timely, and specific. In language testing, *informational* fairness is fostered when the information about the test procedures are thoroughly and reasonably explained, and feedback from the test is provided in a timely manner and specific for each test taker. Adhering to these criteria is considered good practice to ensure fair treatment of test takers. This study elicited whether language learners viewed the information they received during the testing event to be fair and how that perception may have contributed to how they viewed the overall language program.

Research into organizational justice has established that fairness perceptions of the dimensions of justice can predict justice judgments of the organization overall (Blodgett, Hill, & Tax, 1997; Colquitt, 2001). However, this link has yet to be made in the educational context, where research has focused on examining how the dimensions influence student reactions (e.g., Chory-Assad, 2002; Chory-Assad & Paulsel, 2004; Chory, 2007). For example, Chory et al. (2017) reported that when American university students perceived their grades as being unfair given the performance they gave in a course, they were more likely to experience negative emotional responses (disgust, humiliation, stress). In order to relieve this emotional distress, the learners acted out against the instructor of the course (the social entity). They complained to their classmates about the teacher, complained about the teacher to the school administration, and some reported acting verbally aggressive toward the teacher in class. Similar findings were reported by Chory-Assad (2002), who found that university learners who felt the procedures in which grades were assigned were unfair were also less motivated to learn, felt negatively about learning, and acted aggressively towards the instructor. When students felt their instructor treated them with disrespect (*interactionally* unfair), Chory (2007) reported that they also viewed the teacher as being less credible. These studies consistently showed that when students felt they were unfairly assessed, then negative consequences were also reported. It is important to prevent learners from feeling this way, as negative reactions may disrupt the learning process, not only for the individual feeling negatively, but also other learners who may be impacted by his/her actions when negative behaviors are acted out against the instructor. If a teacher spends time generating enthusiasm to learn from an unmotivated student body or addressing behavioral issues in the class, there is less time devoted to achieving instructional objectives and attending to learner educational needs. For the language learning classroom, this can be amplified, since instructional time is limited and the primary source of input in many contexts (e.g., foreign language context) is the instructor and instructional materials used in class. Though the current study does not

address the consequences of assessment unfairness in the classroom, findings from the existing research provide insight into why it is important to examine how learners perceive their assessment events. This has been an under-investigated area of language assessment and is a gap the current study aims to fill.

Conceptual Framework

Though the concepts of fairness and justice have been distinguished from each other, a link between the two has yet to be clearly demonstrated. To establish this connection, the current study aimed to show how perceptions of the fairness of a single testing event related to justice perceptions of the social entity administering the test. The testing event was inclusive of information provided to and communication with learners before the test, during the test administration, and following the test. This study drew from the organizational justice literature, where perceptions of the fairness of a transactional event (e.g., making a complaint) have been found to contribute to perceptions of overall organizational justice (Blodgett et al., 1997; Cropanzano, Byrne, Bobocel, & Rupp, 2001). The conceptual framework for this study stems from this body of literature, as well as the limited number of classroom justice studies demonstrating that fairness perceptions of a single assessment event may also contribute to perceptions of the social entity. As such, it was expected that each dimension of classroom fairness—*distributive*, *procedural*, *interactional*, and *informational*—would share a relationship with one another and with justice judgments of the social entity administering the test. It was also expected for each dimension of classroom fairness to predict perceptions of justice of the language program. Establishing these relationships would provide useful insight into how perceptions of a single testing event contribute to perceptions of the overall entity responsible for delivering that test. Figure 1 illustrates the conceptual framework for classroom fairness of a language test.
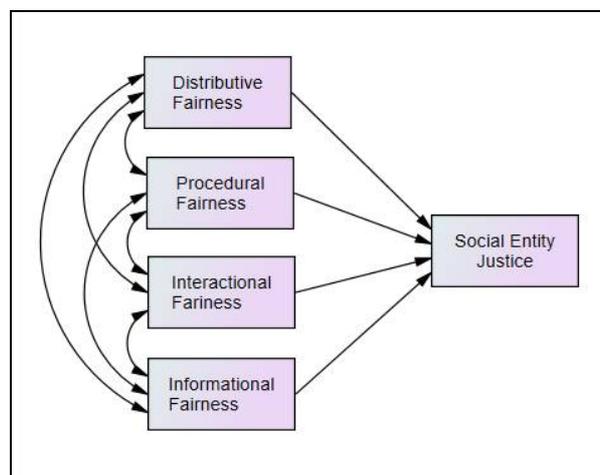

Figure 1. Conceptual framework for fairness and justice in language classroom assessment.

To address the gaps identified in the literature and to test the conceptual framework presented in Figure 1, the following research questions were formulated.

1. How do university second/foreign language learners in Taiwan perceive the dimensions of classroom fairness (distributive, procedural, interactional, informational) of a classroom language test?
2. Are the dimensions of classroom fairness positively related to one another? Are they positively related to justice perceptions of a language program?
3. Do the dimensions of classroom fairness predict justice perceptions of a language program?

Methodology

Participants

A total of 101 students in 14 different universities throughout Taiwan completed an online questionnaire (described below). After eliminating participants who completed the questionnaire improperly (e.g., not indicating they answered the questionnaire based on a classroom-based assessment), responses from 83 respondents were analyzed. Overall, 70% of the respondents were female, 52% of whom were in their first year of undergraduate study. In total, respondents reported studying nine different languages. English (48%) and Chinese (18%) were the most consistently reported language studied, with less than 10% reporting studying Japanese, Korean, French, German, Spanish, Turkish, and Indonesian. Eliciting fairness and justice perceptions from learners in similar contexts (Taiwan university language schools) allowed for observations to be made across languages, and not limited to a single language in a single context.

Instruments and Procedure

An organizational justice survey developed by Colquitt (2001) was adapted for this study to measure the dimensions of classroom assessment fairness—*distributive*, *procedural*, *interactive*, and *informational* fairness—of an individual testing event, as well as the general perceptions of justice of the social entity administering the test. The original questionnaire gauged these dimensions of fairness in the retail business context in the U.S., so the questionnaire was re-worded to fit the needs of the current study in gauging fairness perceptions of a language test administration and justice appraisals of a language program. The original English was translated into Mandarin, and a bi-lingual survey was distributed to the participants in this study.

The questionnaire consisted of 22 statements representing the five intended constructs (see Appendix). Respondents indicated whether their score on a recent classroom language test (e.g., mid-term exam) represented their performance on the test using four items measuring *distributive* fairness. Five *procedural* fairness items elicited perceptions of the test procedures and their equality for all test takers. Four *interactional* fairness items measured perceptions of the respectfulness of communications with the persons delivering the test (i.e., their classroom instructor). Four *informational* fairness items inquired about the perceptions of the quality and quantity of information provided to respondents before and after the test administration. Finally, five *entity* justice items gauged respondent perceptions of the overall justice of the language program (social entity) within which the test was delivered. The order of the items was varied to avoid a possible order effect, and three items were negatively worded to ensure participants responded to each item honestly and avoided marking the same value without reading the statements.

Respondents indicated their agreement with each statement using a five-point Likert scale, ranging from 1 = *strongly disagree* to 5 = *strongly agree*. The standard item alphas of the scales were as follows (see Table 1): *entity* justice (.760), *distributive* fairness (.868), *procedural* fairness (.803), *interactional* fairness (.738), and *informational* fairness (.487). Scores on the entity justice scale ranged from 2.20 to 5.00 (M=3.99, *SD*=.70). Scores from the *distributive* fairness scales ranged from 2.00 to 5.00 (M=3.91, *SD*=.72). Scores from the *procedural* fairness scales ranged from 2.00 to 5.00 (M=4.12, *SD*=.66). Scores from the *interactional* fairness scales ranged from 2.75 to 5.00 (M=4.28, *SD*=.61). Scores from the *informational* fairness scales ranged from 1.75 to 5.00 (M=3.39, *SD*=.63).

Because the questionnaire was adapted for the current study's context, the items for each dimension on the questionnaire were subjected to confirmatory factor analysis (CFA) using SPSS AMOS version 24 to confirm the factor structure of the instrument. Results from the CFA indicated the *entity* justice construct and three of the four dimensions of classroom assessment fairness—*distributive*, *procedural*, and *interactional* fairness—were unidimensional, demonstrating good fit to the confirmatory models. The *informational* fairness items did not demonstrate good fit to the confirmatory model. Due to this poor

model fit and its low reliability estimate, the *informational* fairness scale was removed from the subsequent analysis to answer the research questions.

TABLE 1
Descriptive Statistics, Correlation Matrix, and Reliability Estimates for the Fairness Dimensions and Justice Judgement

	M	SD	1	2	3	4	5
1. Entity justice	3.99	0.70	(.760)				
2. Distributive	3.91	0.72	.582**	(.868)			
3. Procedural	4.12	0.66	.699**	.639**	(.803)		
4. Interactional	4.28	0.61	.670**	.545**	.689**	(.738)	
5. Informational	3.39	0.63	.123	.321**	.344**	.071	(.487)

*Note. *p* < .05; **p* < .01; N = 83; Cronbach's alphas in parentheses

Results

The first research question was interested in how university second/foreign language learners in Taiwan perceive the fairness of a classroom language test and the justice of the language program administering the test. Post-hoc pairwise comparison with Bonferroni correction revealed that participants considered how they were treated by the class instructor during the testing event (*interactional* fairness) to be significantly more fair (M=4.28, SD=.61) than for how equally the procedures used to administer the test were carried out (*procedural* fairness) (M=4.12, SD=.66) and how fairly their score represented their performance on their respective test (*distributive* fairness) (M=3.91, SD=.72). The results also indicated their perceptions of *procedural* fairness were higher than those of *distributive* fairness, suggesting that they thought the procedures were more fairly administered than the score represented their test performance. Overall, the respondents reported feeling their language programs were just institutions (M=3.99, SD=.70).

The second research question inquired about whether the dimensions of classroom fairness were positively related to one another. Results shown in Table 1 indicate that each dimension shared a moderate to strong relationship with one another. *Interactional* fairness and *procedural* fairness had the strongest relationship (*r*=.689, *p*<.01). This correlation estimate was expected, since these two constructs are considered to be closely linked, and past studies (e.g., Chory-Assad & Paulsel, 2004) have combined the two into a single *procedural* fairness factor when examining classroom fairness. *Procedural* fairness and *distributive* fairness also shared a moderate to strong relationship (*r*=.639, *p*<.01). This too was expected, since there is a logical connection between the two—the score awarded and the procedures used to arrive at that score. The weakest correlation among the classroom fairness dimensions was between *interactional* fairness and *distributive* fairness (*r*=.545, *p*<.01). The relationships among the dimensions of classroom fairness suggest they may be correlated, but distinct, aspects of fairness perceptions of a language test.

The second part of research question two asked about the relationship among perceived justice of the language program and each dimension of classroom fairness. Results indicated that all three aspects shared a moderate to strong relationship with justice judgments of the program. The strongest relationship was between *procedural* fairness and program perceptions (*r*=.699, *p*<.01), followed by *interactional* fairness (*r*=.670, *p*<.01), and finally *distributive* fairness (*r*=.582, *p*<.01). These findings indicate that when the respondents felt that the processes through which the classroom test was administered were fairly applied, the communication with the language teacher administering the test were fair and respectful, and when their test score fairly represented the performance they gave on the test, they would also perceive the language program overall as being a just entity. This result provides some empirical evidence that a single testing event can relate to how language learners perceive the justice of their language program.

The third research question inquired about which aspects of classroom fairness predict justice judgments of a language program. The regression model shown in Table 2 was statistically significant $F(3, 79) = 35.33$, $p<.001$. The adjusted R^2 value of 0.56 indicated that 56% of the variance in justice perceptions of the language program was predicted by the three classroom fairness dimensions. The results further show that *procedural* fairness was the strongest significant predictor ($\beta=.365$, $p<.05$) followed by *interactional* fairness ($\beta=.325$, $p<.05$). *Distributive* fairness also predicted justice perceptions but was not a significant predictor ($\beta=.172$, $p=.081$). These results suggest that when the participants in this study formulated a justice judgment about their respective language programs, they considered how fairly the test procedures were applied for all test takers and how fairly and respectfully the classroom instructor administering the test treated them throughout the test administration. However, they did not consider how fairly the score they received on the test represented the effort they gave when completing the test.

TABLE 2
Simultaneous Multiple Regression of Classroom Fairness Dimensions on Justice Judgment of a Language Program

Predictors	β	R	R^2	Adjusted R^2	F
Distributive	.172				
Procedural	.365*				
Interactional	.325*				
Entity justice		.76	.57	.56	35.33

*Note. *$p<.05$; N = 83

Discussion

The results indicate that language learners in this study felt that they were treated fairly by their language instructor during a single test administration. Of the dimensions of classroom fairness the students reflected on, the fairest aspect was the manner in which their respective instructors communicated with them during their test administration (*interactional* fairness). This may be a welcomed result by language teachers in this context, as perceptions of unjust or unfair treatment of students by teachers in other contexts have been associated with perceptions of lower credibility of the instructor. For example, when students in Chory (2007) felt their instructor treated them unfairly, they perceived him as being of low competence, character, and empathy. The author suggested that to avoid the possibility of such negative perceptions of the instructor, and by extension, the educational program, treating learners in a respectful and fair way is essential. This advice is further supported by classroom teachers themselves, whom have identified trust and respect during the learning process (inclusive of assessment) as an important criterion for fair classroom assessment (Tierney, 2014). If students feel they are not treated in a respectful way by their instructors, it could contribute to perceptions of unfairness, and ultimately negative consequences for students and teachers alike. Though the effects of unfair treatment of students by instructors was not investigated in this study, learning how learners perceived the interactions with their instructor during an assessment event would be useful for teachers aiming to avoid such negative reactions.

The results also showed that participants believed that the procedures in which the language tests were administered were fair, although not as fair as the treatment they had received from their instructors. This is a positive indicator for the language programs in the Taiwan universities of this study, as it meets one of the criteria Chinese university students reported as being important for test fairness—that test procedures should be applied equally for all test takers (Song, 2016). By indicating that they felt the procedures were evenly administered in this study, instructors may have avoided negative behavioral responses from the learners. Though not examined in this study, it has been reported that when learners feel the procedures with which they are assessed are applied unfairly, or benefit one group of learners

over another, they can lose motivation to learn, feel less positively towards learning, or act aggressively towards the instructor (Chory-Assad, 2002). However, the results in this study indicate students perceived the procedures to be fairly administered.

The least fair dimension of classroom fairness reported by the participants was *distributive* fairness. The mean score below 4.0 (indicating fair) for this dimension suggests that the participants on average felt their test score representation relative to their performance was almost fair, but not unfair. A noted source of fairness in language testing is adequate construct representation being measured on the test (Kunnan, 2018; McNamara & Ryan, 2011), and if results from statistical analysis indicate construct irrelevant factors may have influenced the test score, then that test is considered to be unfair. It is unfair because the score awarded was not entirely based on the merits of the performance of the intended construct(s), but inclusive of factors extraneous to those constructs. This conceptualization of fairness has been expressed by Chinese university students, who reported willingness to accept test scores as being fair if they are awarded based on merit (Song, 2016), where high scores are given to test takers who perform best. However, if the scores are influenced by irrelevant factors (e.g., having greater content knowledge of the listening/reading texts; having more experience with the test tasks), then the test may be perceived as being unfair; a perception that could lead to learners actively act out against their teachers (Chory et al., 2017). Because the participants in this study did not perceive their scores to be overly fair, it is possible that they felt their scores may have been influenced by some construct-irrelevant factor(s). However, because they reflected on their own language test, it is challenging to verify if each assessment adequately measured its intended constructs, and without follow-up interviews, it is unclear what factor(s) students thought may have influenced these perceptions, or if they did so. Therefore, future studies eliciting such results may aim to examine this.

The moderate to strong relationships among the dimensions of classroom fairness indicate that when the participants felt that the score fairly represented their performance (*distributive* fairness), that they also felt the test procedures were administered fairly (*procedural* fairness) (r=.639, p<.001), and that the instructor communicated with them respectfully throughout the testing even (*interactional* fairness) (r=.545, p<.001). They also felt that when the testing event was *procedurally* fair, it was also *interactionally* fair (r=.689, p<.001). These findings support the conceptual framework of this study suggesting that the dimensions of classroom fairness of a testing event are closely related to one another. When appraising the fairness of a testing event, the perception of one classroom fairness dimension is associated with the perceptions of two others and an increase or decrease in one would be associated with an increase or decrease in another, respectively.

All three dimensions were also moderately to strongly correlated with justice perceptions of the language program. When the participants felt the test procedures were fair, when the score fairly represented their performance, and when the communication among with the instructor was fair, they also considered the language program to be just. This pattern of findings is consistent with the organizational justice literature (e.g., Colquitt, 2001) and provides some evidence of the relationship among the constructs in the educational context. Language programs and instructors teaching within them may benefit from these findings, as they suggest that treating learners respectfully may also relate to how they view the fairness of their score and procedures used to carry out the test, and also how they view the entire program. Maintaining a positive perception towards the teacher and program administering the test is essential to maintaining credibility among the students, as this credibility may be diminished if learners feel they are unfairly treated (Chory, 2007). This perception may persist and ultimately lead to reduced student motivation, negative attitudes towards learning, and negative behaviors (Chory-Assad, 2002). To avoid such negative consequences, it is important to ensure tests are procedurally unbiased, scores are distributed based on merit, and interactions with test administers are respectful.

The results also indicate that language learner perceptions of the fairness of a single testing event are predictive of a justice judgment of a language program. When students felt the procedures were administered fairly and the communication with the classroom instructor was respectful throughout the testing event, they perceived the language program to be just. This pattern of results is partially in line

with Colquitt (2001) and provides partial support for the conceptual framework for classroom fairness in language assessment presented in the literature review. This framework was informed by Cropanzano et al. (2001) from the organizational justice literature showing that appraisal of an event, which in this case was an assessment event based on four dimensions of classroom fairness, predicts justice appraisals of a social entity (the language program). However, the results indicated that only *procedural* and *interactional* fairness did so. As this is the first study to examine the relationship among these constructs in the language assessment context, it provides a foundation from which future studies may aim to expand on the relationship.

The strongest predictor of justice perceptions of the language program was *procedural* fairness. When participants felt the test procedures were administered equally and fairly for all test takers, they perceived the language program overall as being just. This finding was expected given the context in which this study was conducted. Song (2016) reported that Chinese language learners perceived a standardized language test to be fair if its procedures were applied equally for all test takers. Though that study examined student perceptions of fairness of a standardized test context, the findings from the current study conducted in a similar context provide some evidence that they may feel similarly when assessed for a lower-stakes test in the language classroom. Language programs may find this result useful in that their reputations may be viewed positively or negatively based on how students studying within that language program experience a single test administration. If learners feel a test was administered equally and fairly by avoiding bias against/for a group of learners, the program's reputation may have been raised or at least maintained in the eyes of the students. Likewise, language teachers administering the test ought to consider how they administer tests and examine the feedback from students, as attitudes towards these administrations may impact learner perceptions of them and the program within which they work.

Interactional fairness was the second strongest predictor of language program justice. When participants felt they were communicated with respectfully throughout a testing event, they perceived the language program as being more just. This respectful communication included the instructor administering the test refraining from making disparaging remarks to the language learners and treating the test takers with dignity before, during, and after the test was administered. This seems rather intuitive for instructors to do, but it is important to reiterate that the consequences of not doing so may lower perceptions of the teacher and the overall program, as reported in the classroom justice literature (e.g., Chory-Assad, 2002; Chory-Assad & Paulsel, 2004; Chory et al., 2017).

Though all three dimensions shared a relationship with the program justice judgments, *distributive* fairness was not a significant predictor. Here, student perceptions that their test scores fairly represented their performance did not predict how they felt about the justice of the language program. This was a surprising result, considering learners studying in a similar context to the current study (language learning in Chinese university) have reported that when evaluating the fairness of a test administration, it is important for the scores they earn to be based on merit (Song, 2016). However, this may not have extended to justice perceptions of the language program overall in this study, and may have been limited to only that test administration. Similar findings have been reported outside of the language learning context, where *distributive* unfairness shared a moderate relationship with low motivation, negative affect, and aggression towards the instructor, but only *procedural* unfairness predicted all three outcome variables (Chory-Assad, 2002). Chory-Assad (2002) suggested that an interaction effect may explain the result, where *distributive* fairness would have an effect if *procedural* fairness were low. However, this interaction was not found in a follow-up study by Chory-Assad and Paulsel (2004), nor was it found in this study among any of the dimensions of classroom fairness. Despite this, *distributive* fairness not being a significant predictor was unexpected.

Though this study is limited in its scope, these results provide interesting feedback for language programs. They illustrate that a single testing event may influence how language learners view the justice of an entire program. If learners feel test procedures are unequally administered, by favoring or disadvantaging a group of test takers over another on one test, then they may consider the entire program to be unjust. Likewise, if the communication with the language instructor during one testing event is not

considered respectful, then it may reflect poorly on the program. This interpretation should be taken with caution, however, since this design of this study prevents causality from being established. Yet, the empirical results may provide language teachers and language programs with insight into how a single testing event may impact language learner perceptions of them.

Conclusion and Future Directions

Based on the findings of this study, the participants viewed just language programs as those that administer assessments equally for all test takers and treat their learners with respect throughout a testing event. Though preliminary, the findings support an extension to current definitions of fairness and justice in language assessment, one that accounts for the subjective perceptions of the test and social entity administering it. In addition to fairness being an objective test quality verified through statistical analysis by those not directly involved in the testing process, it may also include perceptions of the *procedural*, *interactional*, and *distributive* fairness of a testing event. Likewise, the definition of justice may extend beyond the consequences of test score use, social values present in the test constructs (McNamara & Ryan, 2011), or the qualities of the social entity administering the test and its beneficence to society (Kunnan, 2018) as determined through content analysis, and also include judgments by stakeholders of the entity's justice. The subjective aspects of fairness and justice illustrate how the two constructs are interconnected, with fairness judgments of a testing event (i.e., classroom fairness judgments) predicting justice judgments of the social entity. The existing definitions of the two constructs in the language testing literature are ambiguous regarding how they relate to one another and the conception of the relationship explained here establishes a clear link between the two.

The study is not without its limitations. First, the small sample size and narrow context in which it was conducted prevent the findings from being generalized to the greater language learner population. Future studies are therefore encouraged to include a lager sample size and be carried out in broader L2 learning contexts so that the results may be more generalizable. Further, the study utilized a cross-sectional design preventing causality from being established. To establish the causal relationship among the dimensions of justice and justice judgements of a social entity, future studies should consider utilizing a longitudinal design to examine changes in justice judgments over time based on appraisals of multiple assessments within a course. Finally, though the literature on fairness and justice in language assessment has focused on standardized tests, it may be more worthwhile to continue examining the relationship in the classroom context. Because language assessment mostly occurs in the classroom, it is essential that students feel they are being evaluated fairly. Not doing so may cause them to feel negatively about learning and harm motivation (Chory-Assad, 2002), prompting them to eventually act out against their teacher and the program administering the assessment (Chory et al., 2017). This study represents one of the first to examine fairness and justice in the language classroom context, and future studies are encouraged to explore this relationship further, as well as the effects perceptions of unfairness may have on learner behaviors. Feedback from these studies would provide useful feedback for teachers and language programs alike, and aid them in understanding learner perceptions. With this insight, learning entities may be better able to meet their students' needs and expectations.

Acknowledgements

The Author

Matthew P. Wallace is Assistant Professor at the University of Macau. His research interests cover fairness and justice in language testing, language teaching and learning, and L2 listening comprehension.

Email: mpwallace@umac.mo

References

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267-299): Academic Press.

Bies, R. J., & Moag, J. F. (1986). Interactional justice: Communication criteria of fairness. In R. J. Lewicki, B. H. Sheppard, & M. H. Bazeman (Eds.), *Research on negotiations in organizations* (Vol. 1, pp. 43-55). Greenwich, NY: JAI Press.

Blodgett, J. G., Hill, D. J., & Tax, S. S. (1997). The effects of distributive, procedural, and interactional justice on postcomplaint behavior. *Journal of retailing, 73*(2), 185-210.

Byrne, Z. S., & Cropanzano, R. (2001). The history of organizational justice: The founders speak. In R. Cropanzano (Ed.), *Justice in the workplace: From theory to practice* (Vol. 2, pp. 3-26). London: Lawrence Erlbaum Associates.

Chory-Assad, R. M. (2002). Classroom justice: Perceptions of fairness as a predictor of student motivation, learning, and aggression. *Communication Quarterly, 50*(1), 58-77.

Chory-Assad, R. M., & Paulsel, M. L. (2004). Classroom justice: Student aggression and resistance as reactions to perceived unfairness. *Communication Education, 53*, 253-273. doi:10.1080/0363452042000265189

Chory, R. M. (2007). Enhancing student perceptions of fairness: The relationship between instructor credibility and classroom justice. *Communication Education, 56*, 89-105.

Chory, R. M., Horan, S. M., & Houser, M. L. (2017). Justice in higher education classroom: Students' perceptions of unfairness and responses to instructors. *Innovative Higher Education, 42*(4), 321-336. doi:10.1007/s10755-017-9388-9

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology, 86*(3), 386-400.

Cropanzano, R., Byrne, Z. S., Bobocel, D. R., & Rupp, D. E. (2001). Moral virtues, fairness heuristics, social entities, and other denizens of organizational justice. *Journal of Vocational Behavior, 58*(2), 164-209.

Deutsch, M. (1985). *Distributive justice: A social-psychological perspective*. New Haven, CT: Yale University Press.

Greenberg, J. (1993). The social side of fairness: Interpersonal and informational classes of organizational justice. In R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* (pp. 79-103). Hillsdale, NJ: Erlbaum.

Hatfield, E., Rapson, R. L., & Bensman, L. (2012). Equity theory. In V. S. Ramachandran (Ed.), *Encyclopedia of human behavior* (Vol. 2, pp. 73-78). San Diego: Academic Press.

Karami, H. (2013). The quest for fairness in language testing. *Educational Research and Evaluation, 19*(2), 158-169.

Kunnan, A. J. (2000). Fairness and justice for all. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment* (Vol. 9, pp. 1-14). New York: Cambridge University Press.

Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European year of languages conference papers, Barcelona* (pp. 27-48). Cambridge, UK: Cambridge University Press.

Kunnan, A. J. (2018). *Evaluating language assessments*. New York: Routledge.

Leventhal, G. S. (1980). What should be done with equity theory? In K. J. Gergen, M. S. Greenberg, & R. H. Willis (Eds.), *Social exchange: Advances in theory and research* (pp. 27-55). Boston, MA: Springer US.

Lind, E. A., & Tyler, T. R. (1988). *The social psychology of procedural justice*. New York: Plenum Press.

McNamara, T., & Ryan, K. (2011). Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test. *Language Assessment Quarterly, 8*(2), 161-178. doi:10.1080/15434303.2011.565438

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Education measurement* (3rd ed., pp. 13-103). New York: American Council on Education & Macmillan.

Song, X. (2016). Fairness in educational assessment in China: Historical practices and contemporary challenges. In S. Scott, D. E. Scott, & C. F. Webber (Eds.), *Assessment in education: Implications for leadership* (pp. 67-89). Cham: Springer International Publishing.

Thibaut, J. W., & Walker, L. (1975). *Procedural justice: A psychological analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Tierney, R. D. (2014). Fairness as a multifaceted quality in classroom assessment. *Studies in Educational Evaluation, 43*, 55-69. doi:10.1016/j.stueduc.2013.12.003

Xi, X. (2010). How do we go about investigating test fairness? *Language Testing, 27*(2), 147-170. doi:10.1177/0265532209349465

Appendix

Classroom fairness and justice in L2 assessment questionnaire items.

1. My score reflects the effort I have put into studying for the test.
2. I have been able to express my views about the test procedures if I wanted to.
3. The instructor giving me the test treated me in a polite manner before the test.
4. The test procedures were explained thoroughly before the test.
5. I am able to appeal my score if I want to.
6. My score accurately reflects my performance on the test.
7. Before the test, the explanations of the test procedures (how to complete the test) were reasonable for me to follow.
8. The test procedures have been applied consistently (same for all test takers).
9. The instructor giving me the test treated me with respect during the test administration.
10. The way the language program works is *not* fair.
11. I received feedback about my test performance in a timely manner.
12. I think most of the people taking the test would say they are *unfairly* treated by the language program.
13. Given my performance on the test, my score is justified.
14. The instructor treated me with dignity when I received my test score.
15. For the most part, the language program treats test takers fairly.
16. The test procedures (steps to complete the test) were fair to all test takers.
17. I have received disparaging remarks from the instructor about my test performance.
18. My score is appropriate for the performance I gave on the test.
19. In general, I can count on the language program to be fair.
20. The test procedures were free of bias (i.e., equal for all test takers).
21. The feedback from the instructor about my test performance was tailored specifically for me.
22. Overall, the language program has treated me fairly.