

REVIEW

Open Access



# Reviewing the IELTS speaking test in East Asia: theoretical and practice-based insights

Ethan Douglas Quaid

Correspondence: [ethan.douglasquaid@outlook.com](mailto:ethan.douglasquaid@outlook.com)  
School of Languages, Education and Cultures, The Sino-British College, University of Shanghai for Science and Technology, Shanghai, People's Republic of China

## Abstract

This paper reviews the International English Language Testing System's speaking sub-test in the East Asia region with reference to theoretical and practice-based perspectives and identifies future research opportunities to enhance the measures of test qualities found. The test's construct validity was seen to accurately measure the abilities defined in the IELTS speaking construct; however, high reliability was revealed to the detriment of other test qualities. Conclusions drawn indicate three primary facets of test qualities that could be addressed to increase the IELTS speaking sub-test's usefulness and therefore effectiveness in the East Asian regional context, although these test quality improvements could also be considered as beneficial when applied on a global scale. Firstly, content developers and item writers could provide a greater degree of test item content relevancy to the characteristics of a changing test-taker population. Secondly, multiple future research collaborations between the IELTS partners and institutional test score users seeking to provide better evidence of predictive validity would be beneficial to counteract the lower degree of authenticity shown. And finally, a re-intensification of efforts enhancing positive washback for test takers and exam preparation course providers within the East Asian region is essential.

**Keywords:** IELTS, Speaking test, Oral proficiency interview, Reliability, Usefulness, Test qualities

## Review

### Introduction

The focus of this review is the International English Language Testing System's (IELTS) speaking sub-test (test). The speaking component has been chosen as the basis for this evaluation for two primary reasons. Firstly, the author has been extensively involved with the IELTS oral proficiency interviews (OPI) in the East Asia region and thus able to contribute valuable discussion from an internal and geographical perspective. Secondly, the speaking OPI is salient as it is universally agreed by language testing researchers and professionals that there is difficulty in assessing oral proficiency, because it presents an array of issues which need to be accounted for, addressed, and resolved with regard to their practicality, reliability, and validity. Complicating these issues is that no singular embodying theoretical model has been adept at proving language knowledge and use; therefore, it has been left to the notions of test validity and usefulness to attempt to uphold language testing theory.

### **Background**

In 1980, the English Proficiency Test Battery transformed into the English Language Testing System (ELTS) and represented a marked departure from a traditional structurally focused approach to language testing, by introducing a subject-specific speaking sub-test via individual interview, reflecting communicative language learning. Following the ELTS validation study in 1987, commissioned by Cambridge English and the British Council, international participation was broadened by including the International Development Program Education Australia in the new partnership and the IELTS, in its first form, was implemented in 1989. It was decided that the speaking sub-test should remain, but the content was to be non-specialised in nature. However, very little test procedure was in place to ensure internal consistency and the reliability of test takers' performance, and raters were directed by basic holistic band scales that were necessarily ambiguous.

Throughout the 1990s, only two published studies investigating the test were made available, and findings suggested that the length of different sections of the test showed too much variation. Furthermore, in some sections of the test, examiner talk dominated, because overuse of closed question types to elicit test takers' responses led to a low average of words per turn, which highlighted the need for interview and interlocutor equivalence, and hence fairness to candidates. Boddy (2001) claims that additional unpublished studies commissioned by the IELTS partners were conducted which initiated the 2001 revision of the test.

The objectives of the 2001 revision were to simplify the design, to develop a clearer specification of tasks in terms of input and expected candidate output, to increase standardisation of test management by the introduction of examiner frames, and to revise the rating scale descriptors, which were changed from a holistic global scale to a set of four analytic scales focusing on different aspects of oral proficiency. Taylor and Jones (2001) and Brown (2006) assert this was, in part, because there was a degree of inconsistency in interpreting and applying the holistic band scales. A further revision to the Pronunciation rating scale in 2008 arose as a consequence of the research conducted by Brown (2006) and Brown and Taylor (2006). The four existing pronunciation bands were expanded to nine, mirroring the three other analytic scales. This revision led to the key performance pronunciation features being more clearly specified throughout a greater number of bands on the analytic scale, and subsequently, the criteria became easier for raters to apply.

### **Theoretical models of language knowledge and use**

Through three overarching main approaches of language testing theory: structural, functional, and general proficiency, applied linguists attempt to provide definition to language knowledge and use. The IELTS speaking sub-test, predominantly, is to be found within the general proficiency category, due to it having no syllabus to sample, aiming to establish the standard of English for foreign students wishing to enter universities, and inviting test takers of all language backgrounds and abilities to participate.

However, the implausibility of determining any one of these models as defining, and assuming that users must have all three kinds of knowledge, has led to the test developers using a combination of all three in their quest to determine overall language proficiency. For example, the IELTS speaking band descriptors measure syntactical structure and thus

draw on the structuralist model, whilst task types conform to the functional. Therefore, the test relies on the notion that assessment of speaking seldom conforms to all aspects of a singular theoretical model because the test's truly applied purpose means that concrete features cannot be explained by one theoretical perspective.

In claiming primary use of the general proficiency model, the IELTS speaking sub-test asserts the notion that there is some varying technically analysable, but fundamentally indivisible body of language knowledge within each test taker, and therefore, individuals can be ranked on the basis of this knowledge. Therefore, the test strives to discover proficiency through performance.

Nevertheless, while talking about knowledge, the general proficiency model "is also more orientated towards modelling the process of language use than toward understanding underlying competence" (Spolsky, 1985, p.186), implying that performance over a singular test is rarely able to measure underlying communicative competence. Furthermore, the inability to anchor speaking tests to a singular theoretical model has resulted in the IELTS speaking sub-test development drawing upon numerous models to inform test design and content, and thereby creating a test-specific model. Acceptance that speaking test theoretical models are limited, in that they are only able to inform test development at a fairly abstract level, has initiated a primary reliance on validity and/or usefulness claims to uphold the test as being fit for purpose.

### **Validity and usefulness**

Establishing that the test is theoretically supported is problematic, because validity is essentially locatable on a cline with on-going activity and evidence-based argument determining a test's position to a lower or higher degree on the continuum. Early validity theory asserted there were three superordinate types of validity: construct, content, and criterion-oriented, which then became a central theme for studies of psychological, educational, and language testing. However, Messick (1989, p.20) argued that content- and criterion-related evidence contribute to score meaning, and therefore they should be seen as facets of construct validity.

Messick's unified validity framework, in which different types of evidence contribute in their own way to our understanding of construct validity, fundamentally changed the way in which validity is understood and is now the universally accepted paradigm. Unfortunately, Messick's unitary concept may exclude vital evidence contributing towards this current evaluation, and thus, Bachman and Palmer's (1996) concept of *usefulness* is most appropriate, because it incorporates additional test qualities within the evaluation framework that will be addressed in this review.

### **Reliability**

Reliability, the consistency of measurement and/or the consistency of test performance that is reflected in the test scores, is essential in determining the usefulness of high-stake language tests. A high level of reliability is demonstrated within the test and this is realised through testing procedure and marker variability.

Perhaps the most significant contribution to the test's reliability is from the strict procedural guidelines for examiners conducting the test. For example, examiners are forbidden to deviate from fixed discourse exchanges in part one of the test. By fixing

the frames throughout parts of the test, the opportunity for candidates' performance is predetermined leading to increased reliability. This guidance on what can, or cannot, be said by examiners is not only evidenced in part one, but is present to different degrees throughout the remainder of the test and is the salient contributor to test reliability.

Practice-based insight from China, the largest East Asian IELTS market, suggests that these frame delivery guidelines are mostly adhered to by full-time examiners, yet it is unclear if a greater degree of divergence can be found in other East Asian contexts, where examiners are employed on a part-time basis. Independent research investigating frame divergence by interlocutors is not available. Nonetheless, an IELTS-funded study by O'Sullivan and Lu (2002) found relatively few deviations from the fixed frames in the first two parts of the test, whilst more deviation was noted in part three, although the impact on test-taker output was minimal.

Marker variability in any subjectively scored test is an unavoidable outcome of the nature of the rating procedure. The IELTS (2017) claims that the most recent experimental and generalisable studies based on examiner certification data have shown coefficients of 0.83–0.86 for speaking, which is a relatively high correlation coefficient for a speaking test, although the statistical significance (statistical probability) of this result is not reported, and the studies referenced for the coefficient data are not in the public domain. It is also somewhat noteworthy that independent verification of this claim of scoring reliability or its statistical significance is unavailable. This notwithstanding, the IELTS does dedicate substantial effort to the inter- and intra-rater reliability of the test from a rater perspective.

Firstly, a standard set of professional requirements have been set for the recruitment of all existing and new examiners, ensuring that their qualifications and background are sufficient for the role. Secondly, during initial training leading to certification, as well as bi-yearly recertification, clear guidelines incorporating all of Bachman and Palmer's (1996) suggested procedural recommendations are followed in an effort to minimise intra- and inter-rater scoring variability.

Furthermore, a Professional Support Network (PSN) manages and standardises the examiner cadre. This PSN enables Examiner Trainers to monitor raters on a regular basis using recordings of the OPIs. Moreover, the jagged profile system (identified level of divergence) maintains a further check on both intra- and inter-rater reliability by instigating second-marking of test takers' performances which may have been originally misclassified. Targeted sample monitoring is also conducted with selected test centres providing recorded tests for second-marking by Principal Markers. The rater reliability information obtained from PSN is then fed back into examiner retraining.

A salient disadvantage to the PSN network could lie in the audio-recorded nature of the test events, as paralinguistic features indicating biases are excluded from the standardisation process. A recent study by Nakatsuhara et al. (2017) found differences between audio and video rating of performance in the IELTS speaking test and suggested there are benefits for using the video mode for both rating and standardisation.

The IELTS relies on the combination of the aforementioned measures to typify a high level of reliability within the test. Reliability is a first measurable quality of test usefulness; however, construct validity, the second, is not a necessary condition for the first, whereas

Bachman and Palmer (1996) assert that reliability is a sufficient condition for proving construct validity.

### **Construct validity**

Speaking constructs, as relatively abstract entities, present themselves as problematic to accurately define, which has led to the IELTS language testing community assigning the validation of the construct as primarily an ongoing research activity.

Generally, the IELTS speaking construct is definable as oral proficiency; nevertheless, from an examiner's perspective, this speaking construct is further delineated via the rating criteria which is an operationalized format of the main variables identified within. Oral proficiency is then shown to be reducible to four overarching variables: Fluency and Coherence, Lexical Resource, Grammatical Range and Accuracy, and Pronunciation. These variables are then broken down further to individualised band descriptors. The test score does appear to accurately measure the abilities defined in the IELTS speaking construct; however, more evidence-based argument is needed to ascertain if the speaking construct measures all the abilities required of the Target Language Use (TLU) domain.

Convergent validity can be achieved by relating the speaking sub-test score to the test takers' scores generated by the other three skills' sub-tests: Writing, Reading, and Listening. However, attempting to establish a degree of convergent validity presents the IELTS with the problematic issue of showing high correlation between sub-tests, although the constructs, traits, and skills identified as being tested are markedly different. However, ensuring these sub-test inter-correlations are not too low, prevents identifying the items or tasks throughout the entire test as being less homogenous, and thus having a lower internal consistency. Convergent validity allows the test developers to argue for construct validity, because more than one trait is not identified as being measured.

Predictive validity is shared equally between test developers and the institutional test score users, because entry requirements are set by the user. Predictive validity has been a research area within the ongoing validation of the speaking sub-test which has thus far proved elusive to definitive research findings. For example, in a study consisting of two parts, initially, Ingram and Bayliss (2007) found that participants were generally able to produce, in the context of their academic studies, the language behaviour predicted by an IELTS test score, whilst in contrast, Paul (2007, p.2) concluded that 'language production at a micro level similar to that in the IELTS tasks is not necessarily an indicator of overall language adequacy at a macro level or successful task completion'. A focus on studies supporting predictive validity would be enormously helpful in further validation of the tests' construct validity. Central to this research is test content relevance and coverage, encompassing the authenticity of task types and characteristics.

### **Authenticity**

The test, with the aim of demonstrating general proficiency, strives to demonstrate that performance is generalisable to real-life domains, where language is used essentially for purposes of communication. The IELTS (2017) asserts that the test 'is interactive and as close to a real-life situation as a test can get', hereby acknowledging that although the test is administered via an OPI, which contains direct test tasks, it is still nonetheless an indirect

performance-referenced test. However, although the overall argument reinforces the test's face validity in the eyes of most, it is upheld to a lesser degree when analysed more closely.

Perhaps the most notable negation to the IELTS claim of an interactive and close to real-life speaking test is that the discourse is strictly controlled by the interlocutor. Much like Sinclair and Coulthard's (1975) classroom IRF discourse analysis model, both parts one and three of the speaking sub-test exhibit strong evidence of primarily eliciting three part exchanges, as the following analysed part three sample excerpt illustrates.

*Interlocutor:* You don't think of it as a healthy way of thinking? (*Initiation*)

*Candidate:* It's probably not honest to yourself. You can understand what I mean?

(*Response*)

*Interlocutor:* Yes. (*Feedback*) And do you think this will change? (*Initiation*)

(adapted from IELTS, 2017)

This interactional patterning shares similarities with some sub-varieties of L2 classroom discourse and with that found in universities; however, it is very different to everyday ordinary conversation (Seedhouse & Egbert, 2004). This type of institutional discourse does contribute towards standardisation and reliability but has not been proven as an indicator of predictive validity for test takers' future academic or professional success as yet, so its inclusion is purely speculative.

The institutional discourse elicited from the positions and roles adopted in the relationships between interlocutors and test takers suggests that an IELTS OPI consisting of paired test takers may approximate to 'real-life' conversation better. A paired format can deemphasize the goals that are relevant to the test event, lift constraints on the quality and quantity of test-taker output, and encourage a range of behaviour that is atypical of institutional discourse.

Brooks (2009) found that test takers performed better in a paired speaking test than when they interacted with an interlocutor only. This increase in performance was attributable to the co-construction of test discourse which resulted in more linguistically complex performances. Furthermore, the discourse showed more interaction and negotiation of meaning.

Nevertheless, paired format OPIs introduce a number of other problematic factors that relate to construct definition, reliability, and fairness. These facets can include language proficiency (Bonk & Van Moere, 2002), personality (Berry, 2007), and gender (O'Sullivan, 2000) and familiarity (O'Sullivan, 2002). Variables relating to the socio-cultural could be especially important for high-stake tests that are administered globally, as regions such as East Asia have markedly different norms from other contexts.

Unresolved limitations may exist precluding the use of a paired test-taker format in the IELTS speaking test. Nonetheless, if the type of discourse produced during the majority of the test is relatively statutory, this affects the degree to which test takers' characteristics are engaged and subsequently test interactivity.

### Interactiveness

Test taker characteristics, such as language ability, background knowledge and motivations, are the personal attributes of the candidature, whilst interactiveness is the extent of the involvement of these characteristics in completing the test tasks using the same capacities they would use in the TLU domain. As well as predominantly containing institutional discourse features which do not engage and measure the test takers' cognitive processes required to interact in ordinary everyday conversation leading to construct irrelevant score variance according to individuals' physiological, psychological, or experiential characteristics, the IELTS speaking tasks and format can exhibit bias against interrelated sets of characteristics such as test takers' age (physiological), topic knowledge and world view (experiential), or an individual's characteristic (psychological).

Adolescents at increasingly younger ages and in larger numbers are choosing to complete the IELTS test for educational and immigration purposes. For example, Canada has requested IELTS scores for international migrants as young as 14 (Government of Canada, 2017). Topic knowledge and the way in which younger test takers view the world (experiential characteristic) is markedly different from older candidates. Therefore, a business-related test task may suffer from a degree of inappropriateness. It is imperative that test designers account for changes in the test-taker population and respond accordingly, by adapting or phasing out unsuitable test items and tasks.

Rote learning is a psychological characteristic that is prevalent in East Asian educational contexts and its use could be attributed to the nature of the assessment systems and practices in place (Wong, 2004). The institutional discourse mainly consisting of providing information through declarative speech acts, in addition to part two's storytelling task type, realised through an individual long turn, encourages inclined test takers to attempt memorising large amounts of text. Although test takers' responses to this task may belie their true proficiency, the discourse is likely to be tangential to the task prompt and thus not meet the desired level of communicative competence expected (Park & Bredlau, 2014).

As proving memorisation is fraught with difficulty and performance is assessed throughout all parts of the test, East Asian test takers attempt to improve their performance through misuse of this characteristic. IELTS examiners are instructed to rate test takers' performance across all parts of the test which may encourage test takers to memorise part two's long turn responses, even though part three's less standardised format can often provide examiners with a better indication of true performance and score. Perhaps, the most salient question is whether this memorisation washback can be attributed to test impact, test-taker characteristics, specific rote learning methodological contexts, or indeed other forces within the wider educational scene.

### Impact

The IELTS training courses English language instructors have previously taught may to some degree have demanded *teaching to the test*, in preference to seeking an overall improvement in students' language proficiency. This tension between pedagogical and ethical practice is realised through teachers narrowing their instruction to meet what they perceive are the demands of the test construct.

Assigning responsibility for this washback is not as easy as it may appear. Although instructors are perceived to hold a central role, which has led to studies which have

primarily investigated how test washback influences classroom teachers, preparation material developers and learning institutions and their influence on cognitive strategies implemented by test takers are also essential contributory factors. The IELTS speaking sub-test can certainly be said to influence all parties discussed through its rigid test format, serving to increase reliability, and the bi-annually recycled content aiding practicality.

### **Practicality**

Bi-annual production of material can be said to aid the test's practicality, yet it also presents the need for increased test security by way of secure storage of test material, and additional procedural tasks, which require additional administration to complete. However, the most impractical aspect of the test is the continued use of an OPI, using perhaps unnecessary human, material, and time resources, even though the computer-based (CB) IELTS has begun to offer the Listening, Reading, and Writing tests.

Studies including those by Rosenfeld et al., (2005) and Mousavi (2009) have found that comparisons of test-taker scores between CB-OPIs and direct OPIs correlate highly, signalling the possibility that no adverse effects on reliability are to be found, and test takers have a highly positive attitude towards a CB speaking test compared to an OPI of speaking proficiency, due to the CB mode of delivery being more comfortable and less threatening. The overwhelming reason for the IELTS continued support for the OPI has been face validity; however, test takers' perceptions may be changing.

### **Conclusions**

The IELTS speaking sub-test displays strong evidence of high reliability. Contributing towards this are the procedural guidelines for administration of the test and low marker variability. The test score does appear to accurately measure the abilities defined in the IELTS speaking construct, although ongoing evidence-based debate is required to ensure the speaking construct measures all the abilities required in the TLU domain. Additional research studies in this area and of the test's predictive ability would add support to the construct validity argument made.

The test exhibits a lower degree of authenticity, because of the relatively statutory institutional discourse of the OPI helping to increase reliability. This restriction in authenticity also affects test interactivenss by not engaging and measuring the cognitive processes required to interact in ordinary everyday conversation and test-taker characteristics indicative of the TLU domain. Two additional challenges to test interactivenss are identified. Firstly, the changing test population dictates that task content will need to be periodically reviewed in order to fully engage the physical characteristics of ever increasing numbers of younger test takers.

And secondly, East Asian learners with prolonged exposure to rote learning methodology are presently unwittingly encouraged to use this as a test-taking strategy, due to the test's rigid institutional discourse serving to increase reliability. Reliability also affects test impact by encouraging negative washback in exam preparation classes; however, test developers are not the only participants encouraging *teaching to the test*. Equal responsibility needs to be taken by preparation material developers, learning institutions, teachers, and the test takers themselves.



**Acknowledgements**

No acknowledgements to be made.

**Funding**

No funding was received during the process of writing this review article.

**Availability of data and materials**

This is a review article therefore no data will be shared.

**Authors' contributions**

The author read and approved the final manuscript.

**Competing interests**

The author declares that he has no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 30 October 2017 Accepted: 17 January 2018

Published online: 06 March 2018

**References**

- Bachman, LF, & Palmer, AS (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Berry, V (2007). *Personality differences and oral test performance*. Frankfurt: Peter Lang.
- Boddy, N. (2001). The revision of the IELTS speaking test. *Shiken: JALT Testing & Evaluation SIG Newsletter*, 5(2), 2–5.
- Bonk, WJ, & Van Moere, A (2002, December). *L2 group oral testing: The influence of shyness/extrovertedness and the proficiency levels of other group members on individuals' ratings*. Singapore: Paper presented at the AILA Congress.
- Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing*, 26(3), 341–366.
- Brown, A. (2006). An examination of the rating process in the revised IELTS speaking test. *IELTS Research Reports*, 6, 1–30.
- Brown, A, & Taylor, L. (2006). A worldwide survey of examiners' views and experience of the revised IELTS speaking test. *IELTS Research Notes*, 26, 14–18.
- Government of Canada. (2017) *Citizenship Act with Bill C-6 Amendments*. Retrieved January 25, 2018, from [https://www.canada.ca/en/immigration-refugees-citizenship/news/2017/06/bill\\_c-6\\_receivesroyalassent0.html](https://www.canada.ca/en/immigration-refugees-citizenship/news/2017/06/bill_c-6_receivesroyalassent0.html)
- IELTS (2017). *IELTS*. Available Nov 12, 2017, from <https://www.ielts.org/>
- Ingram, D, & Bayliss, A. (2007). IELTS as a predictor of academic language performance – Part 1. *IELTS Research Notes*, 7 Retrieved from [https://www.ielts.org/-/media/research-reports/ielts\\_rr\\_volume07\\_report3.ashx](https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx).
- Messick, S (1989). Validity. In RL Linn (Ed.), *Educational measurement*, (pp. 13–103). New York: Macmillan/American Council on Education.
- Mousavi, SA. (2009). Multimedia as a test method facet in oral proficiency tests. *International Journal of Pedagogies & Learning*, 5(1), 37–48.
- Nakatsuhara, F., Inoue, C. & Taylor, L. (2017). An investigation into double-marking methods: comparing live, audio and video rating of performance on the IELTS Speaking Test. *IELTS Research Reports*. Retrieved from [https://www.ielts.org/-/media/research-reports/ielts\\_online\\_rr\\_2017-1.ashx](https://www.ielts.org/-/media/research-reports/ielts_online_rr_2017-1.ashx)
- O'Sullivan, B. & Lu, Y. (2002) The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. *IELTS Research Reports*. 6.4. Retrieved from [https://www.ielts.org/-/media/research-reports/ielts\\_rr\\_volume06](https://www.ielts.org/-/media/research-reports/ielts_rr_volume06)
- O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System*, 28(3), 373–386.
- O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing*, 19(3), 277–295.
- Park, E. & Bredlau, E. (2014) Expanding the Question Formats of the TOEIC® Speaking Test. *Education Testing Service*. Retrieved from <https://www.ets.org/Media/Research/pdf/Expanding%20the%20Question%20Formats%20of%20the%20TOEIC%20Speaking%20Test.pdf>
- Paul, A. (2007). IELTS as a predictor of academic language performance - part 2. *IELTS Research Notes*. 7.4. Retrieved from [https://www.ielts.org/-/media/research-reports/ielts\\_rr\\_volume07\\_report3.ashx](https://www.ielts.org/-/media/research-reports/ielts_rr_volume07_report3.ashx)
- Rosenfeld, E, Bernstein, J, Balogh, J. (2005). Validation of an automatic measurement of Spanish speaking proficiency. *Acoustical Society of America*, 117(4), 2428–2428.
- Seedhouse, P. & Egbert, M. (2004). *The interactional organisation of the IELTS Speaking Test*. Retrieved from [https://www.ielts.org/-/media/research-reports/ielts\\_rr\\_volume06\\_report6.ashx](https://www.ielts.org/-/media/research-reports/ielts_rr_volume06_report6.ashx)
- Sinclair, J, & Coulthard, RM (1975). *Towards an analysis of discourse*. Oxford: Oxford University Press.
- Spolsky, B. (1985). What does it mean to know how to use a language? *Language Testing*, 2(2), 180–191.
- Taylor, L, & Jones, N (2001). *University of Cambridge Local Examinations Syndicate Research notes 4*. Cambridge: University of Cambridge Local Examinations Syndicate.
- Wong, JK. (2004). Are the learning styles of Asian international students culturally or contextually based? *International Education Journal: Comparative Perspectives*, 4(4), 154–166.