



SETs (Student Evaluation of Teaching): Lessons to learn from differing biases of Chinese and UK students

Neil Allison¹

Abstract

Student Evaluations of Teaching (SETs) have become somewhat controversial tools for involving students in improving the education process, promoting quality assurance, and providing metrics to compare courses and teaching. They are thus, unsurprisingly, one of the most researched areas of higher education. Among small cohorts with a significant variety of cultural backgrounds, validity of evaluation tools may be confounded. This mixed-methods case study took place in the context of foreign language learning at a UK university with 117 language learner survey participants, seven also participating in a diary exercise, to explore a potential confounding variable on SETs in survey form - that of student cultural background. Student attitudes and biases were examined in the light of previous research showing differences between Eastern and Western psychological biases. Results in this educational context support a clear cultural difference between two groups that were compared in this study: Chinese and UK students. Key differences identified were in participants' choice to focus on positive or negative experiences and to what they attribute these experiences. Implications for incorporating SETs as part of course and lesson design are outlined.

Keywords

student evaluations of teaching, foreign language learning, cultural difference

¹ University of Glasgow, Glasgow, UK

Corresponding Author:

Neil Allison, Hetherington Building, University of Glasgow, Glasgow, G12 8RS, UK
Email: Neil.Allison@glasgow.ac.uk

Introduction

Quality and value are important for higher education providers, particularly in a context of increasing student fees and competition, and there is, as part of this, a need to provide metrics to compare institutions (Angell et al., 2008; Sultan & Wong, 2018; Woodall et al., 2014). While this quality relates to a variety of university services, teaching is clearly a critically important service for overall student satisfaction with their university experience (Langan & Harris, 2019), and student satisfaction is linked to loyalty and customer recommendations, as marketing would label them (Athiyaman, 1997); thus measures of teaching quality have become more prevalent and research on them similarly prevalent (Sharpe, 2019).

“Teaching excellence”, it is widely agreed, is complex and as a result is evaluated through various means in order to take account of teaching quality, the learning environment and so on; however, use of student surveys is always at least one method utilised in a UK context (Gunn, 2018). Across the world, assessing teaching quality via Student Evaluations of Teaching (SETs) i.e. in a form that centres on student opinions and perceptions, particularly in survey form, is a common way formative and summative feedback is gathered on teaching quality (Spooren et al., 2013). Despite the large amount of research on such approaches to evaluating teaching quality, examination of cultural differences that might affect responses or attitudes to SETs have been absent. In small classes dominated by a particular national group, any influence of national, cultural, or linguistic factors could be significant. An underlying difference in student understanding of prompts used to gain feedback or differences in attribution of causes of educational success and failure would bring into question how SETs should be administered with small classes or cohorts e.g. how they should be communicated and explained, and how they should be interpreted. These concerns are of particular importance in the 21st century with international students making up an extremely important market for universities in English-speaking nations such as Scotland, England, USA, and Australia, and with Chinese students making up a large proportion of these students (ICEF Monitor, 2022).

Understanding more about international students’, especially Chinese students’, perceptions of SETs and comparing how these may differ with those from a Western nationality thus seems timely. Exploring the validity of SET tools in this context may improve their value for educational institutions and students. For institutions, do high numbers of particular nationalities influence results in SETs? For students, if SETs aim to empower them to have a voice in their education (Pineda & Seidenschnur, 2021), is this evident from students from specific nationalities? This mixed-methods case study took place within the context of second language study at a Scottish university in order to explore these questions in depth. Although this research is small scale and takes place in the context of language teaching, findings should be of interest for those leading small group teaching (any classes of under 25 students).

This research focuses on prompts in SETs that occur in a survey format. Other means of gaining student feedback may introduce different variables, may allow for dialogue and more complex language, and thus separate research in such contexts would be worthwhile.

What are SETs

SETs are any means used to gain feedback on aspects of students' experience of their education, especially regarding the quality of teaching, whether formative (in-course) or summative (end of course). SETs in survey form, as a means to understand the student education experience, began in the 1950s in the US, and were becoming increasingly systematised by the 1970s (Pineda & Seidenschnur, 2021), with marketing and competition as the dominant framework from which to consider that student experience i.e. thinking of the student as a consumer; as someone whose satisfaction is important to the ability of an institution to compete in the market (Langan & Harris, 2019). However, SETs have not only had such a marketing-informed formulation; SETs have also long had as their aims the improvement of teaching and assurance of quality (Chalmers & Hunt, 2016; Marsh, 1984).

For illustration in the UK context, consider the well-known SET the National Student Survey (NSS). This started in 2005, "to inform prospective student choice", "enhance student experience" and "contribute to public accountability" (Diamond et al., 2015, p. 3). The results of the NSS feed into league tables (OfS, 2023a) and into the teaching excellence framework (OfS, 2023b). University teachers in the UK will be familiar with the NSS and perhaps other policies such as the Student Learning Experience model that provide guidance to institutions in Scotland on various ways to involve students in quality enhancement and assurance (Sparqs, 2023). Meanwhile, across the globe there will be few who are not familiar with means of seeking feedback from students on their education experiences, even if these are of their own informal design within their classrooms. Classroom Assessment Techniques (CATs), articulated by Angelo and Cross in their popular handbook from 1993, cover the same territory as SETs in their section 'Assessing Student Reactions to Teachers and Teaching Methods, Course Materials, Activities, and Assignments'. An example of a prompt from Angelo and Cross is "How helpful are my comments on your papers and responses to your written questions in helping you understand the course material?" (Angelo & Cross, 1993, p. 139). An updated version of this handbook looks at these under the label "perceptions" (of students) (Angelo & Zakrajsek, 2024).

Although there have been many updates to guidance on how students should be involved in feedback on their experience, the dominant lens globally for SET surveys remains satisfaction (Spooren et al., 2013) via subjective student perception prompts e.g. satisfied; like; enjoy; engaged (Stroebe, 2020). This use of satisfaction to measure quality in education applies not just to Western universities. For example, satisfaction is one of five benchmarks for quality evaluation used by China's Ministry of Education (Luo et al., 2019). A way of expressing the link between satisfaction and quality is that satisfaction is a surrogate for quality (Langan & Harris, 2019) or a proxy (Harvey & Green, 1993), meaning quality can be indirectly measured by measuring satisfaction.

Literature on satisfaction has long stressed fitness for purpose (comparing outcomes to aims, though taken from the point of view of the service users, not from other stakeholders' points of view) as a key component in forming the link between satisfaction and quality; an example would be the SERVQUAL model in marketing where quality is the gap between perceived performance/outcome and expected performance (i.e. aims) (meeting expectations is good, positive disconfirmation is good, negative disconfirmation is bad) (Sultan & Yin Wong, 2012). Such satisfaction models assume or accept that satisfaction (and dissatisfaction) is an internal state akin to attitude and that it is

transaction-specific, meaning that satisfaction/dissatisfaction is the result of the evaluation of a specific transaction or consumption experience (possibly being linked to expectation/disconfirmation as in the SERVQUAL model). In other words, satisfaction is a subjective construct based on expectations and confirmation or disconfirmation once something has been experienced. This idea of quality being satisfaction and satisfaction being heavily influenced by fitness for purpose ties clearly to the notion of the student as customer (Harvey & Green, 1993) and meeting expectations (Prakash, 2018).

SETs in the context of language classrooms

One problem when trying to understand quality teaching is that different contexts affect learning and influence students' perceptions of their experience of learning (Ghedini & Aquario, 2008), with different disciplines seeing teaching quality in different ways (Gunn, 2018). In the context of language learning, it is quite clear that learning cannot be calibrated or measured in the same way as with subjects primarily focused on the development of declarative knowledge, for one thing because language development is highly complex and not always predictable (see for example, Ellis, 2012, Chapter 6). Effective language teaching methodology involves an array of teacher choices. The most essential and specific to language teaching is the form of environmental stimulation for learners' cognitive processes; teachers need to provide a context for input (on vocabulary, discourse, and so on), design tasks for output to improve speaking and writing, stimulate memory processes and grade materials and tasks suitably for student levels (Pawlak, in Drożdżał-Szelest et al., 2013), complicated when foreign language learning is often done in the target language (Brown, 2009).

Differing beliefs between educator and student are noted as being particularly stark in language teaching (Loewen et al., 2009). For example, the popular communicative approach in language teaching methodology (less presentation on grammar forms; more emphasis on realistic contexts to practise; emphasis on meaning and realistic language; maximising of co-operation and interaction) often does not chime with student preferences (or to use the language of service quality, is a disconfirmation) (Brown, 2009). Preferences and general beliefs about effective learning may well have cultural tendencies (Duisembekova & Özmen, 2020). Satisfaction may be damaged where expectations differ between teacher and student i.e. a negative disconfirmation. Thus, "teachers may need to help students understand some empirically proven principles of second language learning (e.g., the importance of output, interaction, and negotiation of meaning)" (Schulz, 1996, as cited in Brown, 2009, p. 54).

All of the above exacerbate a general problem with SETs, namely that SET scores very often do not correlate with student learning, resulting in question marks over SET validity (Clayson, 2009; Stroebe, 2020). Zabaleta (2007), conducting research in language teaching, found a moderate correlation between poor student performance and low feedback scores but no correlation between high performance and high evaluation scores. He also found inexperienced teachers – Graduate Teaching Assistants ("GTAs", often PhD students who are very new to teaching) – received higher scores than experienced colleagues even though one might expect quality to increase with training and experience (Zabaleta 2007).

Cultural biases affecting SET validity

SETs that focus on student satisfaction (and other similar subjective perceptions – engagement, preference, and so on) as a surrogate for quality are problematic for reasons

other than those highlighted above on language learning (complex learning curve; disconfirmation of expectations). This is the problem of biases, where perceptions such as satisfaction are influenced by factors not connected to teaching quality. Biases that are claimed to affect the validity of SETs include those that are sexist (Sinclair & Kunda, 2000) racist (Heffernan, 2022), and directed against non-native English speakers (Onweugbuzie et al., 2007, cited in Sharpe, 2019, in respect of a study in the USA).

In language teaching in higher education, much is done via smaller groups (typically fewer than 20 students), in contrast to lecture format, to enable more opportunity for student-to-student and student-to-teacher communication. Of course small samples pose significant challenges for evaluating teaching quality through simple survey-based SETs such as the NSS, with increased risks of individual biases skewing overall results; Feistauer & Richter's (2017) work is particularly pertinent here: while they support the belief among SET proponents that the teacher and teaching quality are major sources of variance in SETs, they argue that students account for almost as much variance as teachers (partly due to effects of student personality traits) and propose that teaching quality can be assessed reliably via average scores only if the sample is big enough (probably at least 25 students) (Feistauer & Richter, 2017).

However, even if we have large samples, what if we are comparing SET scores for courses or classes dominated by Western students with those dominated by Eastern students? A student factor that may affect validity is culture, certainly as regards East and West e.g. China and UK in the context of satisfaction in education (see for example Li et al, 2018; Mavondo et al., 2004). Interestingly, certain culturally specific bias differences between East and West that have been observed in psychology literature have not been adequately considered in SET contexts. Specifically, the psychological construct known as attribution bias has not been the focus of studies in the use of satisfaction measures to evaluate teaching quality.

Attribution concerns causal explanations for experiences. An example is where one might explain to oneself a failure in an exam as being caused by poor teaching (Graham, 1991). Two types of Western bias of particular interest to the teaching context coalesce under this broader attribution framework, both of which are supported by psychological research: fundamental attribution error and self-serving/hedonistic bias. Fundamental attribution error is a form of bias involving the tendency to overestimate the personal in situations i.e. the personal characteristics in others for situations that affect us (Ross & Nisbett, 2011). Self-serving attribution bias/hedonistic bias, meanwhile, applies in situations that require a reflection on a positive performance such as in educational performance where we are more likely to attribute causes of such positive events to something internal; by contrast, we will attribute negative outcomes to something external (Shepperd et al., 2008). Research in social psychology evidences a difference in the "construal of the self" between East and West meaning that in the East there is far less pronounced fundamental and hedonistic biases; in the East there is a more holistic or interdependent perspective compared to the West (Lewis et. al, 2008), sometimes referred to as "collectivist" (Maddux & Yuki, 2006). The interest here for SET research is the possible impact a different cultural bias might have on how students perceive their education experiences and the causes of their experiences.

Research gap and aims

Although student evaluations of teaching effectiveness is one of the most thoroughly studied topics in higher education literature, higher education foreign language teaching with small classes and mixed nationalities is sufficiently distinct to warrant separate investigation in order to assess the value of student feedback and to consider improvements. There is little research on cultural differences in attitudes and biases and their possible effects on validity of evaluation prompts in SET tools. Some observations have been made regarding Chinese students using Western-centric SET approaches and their attributions, though not with reference to hedonistic bias (Pratt et al., 1999). This gap is particularly important to explore for educationalists dealing with small class sizes as in typical language classes and when there are high percentages of Eastern students or Western students. Sharpe (2019, p. 39) comments that “Interpretation of data must also be based on an understanding of the sample who completed it”. With increasing representation on courses of Chinese students, differences in attribution could significantly affect evaluations of the educational experience.

Research questions

This study sought to explore if cultural differences between “East” and “West” in attitudes and biases affect students’ engagement with and responses to SET surveys?

The following sub-questions arose, with Chinese becoming the focus for “East” and UK becoming the focus for “West”:

1. To what extent do Chinese and UK students value SETs, “value” being indicated by whether they:
 - a. believe they fill them in honestly, and
 - b. believe they lead to course improvement;
2. Do Chinese and UK students focus on different attributions (causes) and different loci of effect (e.g. themselves, the class) when presented with various SET prompts:
 - a. Satisfied
 - b. Attentive
 - c. Good quality
 - d. Engaged

Methodology and methods

A mixed methods approach was taken. The methodology was influenced by phenomenology with the aim to obtain information about participants’ experience within their social context – language learning at university (Cibangu & Hepworth, 2016) - taking account of interpretivist, exploratory, and positivistic perspectives (Arthur et al., 2012). The view of mixed methods here is in line with the idea that we are not seeking increased validity but hope to improve our understanding of a complex social phenomenon (Sale et al., 2002). Data collection involved a survey, together with a diary/log completed by a small sub-sample of participants, and a focus group with the same sub-sample to aid the interpretation of these data points.

Survey

The survey aimed to capture participants' experiences of SETs, such as their feelings on giving feedback to teachers and their perceptions of feedback's purpose.

The survey contained direct measures of attitudes linked to feelings on giving feedback and students' perceptions of SETs' purposes, but also used indirect measures (Danioni et al., 2020) to understand more about participants' mental associations of key SET prompts such as "satisfied". Indirect measures were predominantly via word-picture associations (participants are asked to choose a picture to associate with a cue word/phrase or "stimulus") and open-ended word association tasks (participants are given a stimulus and they supply their own word that they associate with the stimulus). There are differing terms used in connection with this approach. The origin of this approach in psychology broadly comes from experimental techniques used to reveal unconscious attitudes, first associated with Jung (e.g. Jung, 1910). Such approaches have also been called semantic association as part of semantic network activation in linguistics and cognitive psychology (e.g. Playfoot et al., 2018), while the term "projective techniques" has been used in marketing and consumer research where such techniques are believed to be less influenced by response bias (Mesías & Escribano, 2018).

An additional underlying theoretical influence on word-association tasks was linguistic relativity. This is the theory derived from psychology, anthropology, linguistics, and philosophy (Hoffman et al., 1986) that helps to explain differing mental constructs and the differing effects of the same stimuli based on language and culture: "all observers are not led by the same physical evidence to the same picture of the universe, unless their linguistic backgrounds are similar, or can in some way be calibrated" (Whorf, 1956, cited in Hoffman et al., 1986, p. 1097). Under one linguistic relativity hypothesis, language acts as a spotlight whereby different languages encourage users to attend to different aspects of experiences (Wolff & Holmes, 2011). Word association and picture association have also been used in linguistic relativity research (Wolff & Holmes, 2011) in addition to the aforementioned research within psychology and marketing.

The survey was reviewed by two experienced education researchers and then piloted with three language teachers and two language students: one Chinese, and one from the UK. Instructions for tasks were adjusted in light of feedback prior to sharing with participants.

Diaries

To gain further insights into participants experiences with SET surveys, and recognising the intentionality associated with a phenomenological paradigm (Davidsen, 2013), participants were provided the opportunity to complete a diary exercise based on prompts almost identical to those they would be familiar with as they are used at the university in a survey SET e.g. "I was satisfied with the quality of the class". The word "because" was added to invite information that would be revealing of attributions; in other words, what caused their satisfaction and so on. The use of diaries as data collection tools can support interpretative phenomenological research and enable personal and cultural expression (Willig, 2013). Diaries enabled discourse and function to interact with key concepts under investigation (in the context of this study concepts of satisfaction, quality, engagement; and attention) as these may influence meanings of concepts (concepts may be discourse context dependent (Mayring, 2014, p. 39). Diaries were logs of student experiences after language classes. An event-based protocol was used (Iida et al., 2012) to gather student reflections on lessons participants had attended.

Participants in the diary exercises were all studying on the same programme (participants who volunteered were all Chinese students studying on an intensive English language course having 10 x 1.5-hour lessons per week). With this being the case, the researcher was able to use the timetable for this programme to identify classes where students should provide diary entries and then include this in the instructions. Participants were asked to choose a pseudonym so that it was not possible to link diary responses to particular individuals and particular classes, their diary entries being collected via a form. For ethical purposes, this also ensured it would be impossible to identify the teachers.

The event-based protocol set out the following prompts:

1. I was/wasn't satisfied with the quality of the lesson and teaching because ...
2. I was/wasn't engaged in the lesson because ...
3. The tutor explained/didn't explain things well because ...
4. I was attentive/paid attention OR I wasn't attentive/didn't pay attention because ...

These prompts were chosen because they are the same as those commonly used in a SET employed by their university while prompt number 4 was provided because it has been used for measuring attention in the classroom and seemed relevant to prompt reflection on classroom experiences (e.g. Allison, 2020). Participants were instructed to write an entry for each of these prompts on two separate occasions i.e. for two separate classes, aiming to produce eight log entries in total.

Ethical approval was gained prior to commencing data collection. The researcher had no relationship to students surveyed and student anonymity was maintained in data collection and processing. In terms of positionality of the researcher, he is a language teacher with experience of teaching students from China and from the UK and applying survey-type SETs required in his institution and also experience using his own bespoke in-course and end-of-course tools to understand student experience. This research was borne out of an interest in understanding more about what students think about when completing SETs. On completing the literature review on hedonistic bias, the research became more focused on comparing "Eastern" and "Western" students to explore whether SETs revealed the same biases in SET contexts as in the contexts previously reported in psychology literature.

Data collection

A purposive sample was sought by emailing language students of the university's languages school. Participants were requested to complete a survey and provided the opportunity to participate in a diary exercise and focus group. Participation in the student survey were self-selecting, and no data was specifically sought concerning language studied or level of each participant. Participants were informed as to the nature of the study via a participant information sheet with consent sought through completion of consent forms.

Participants were asked to identify their nationality through a free text question in the survey. In order to compare cultural differences, comparisons were made between UK and Chinese participants. Any participants outside of these two nationalities are referred to as "other".

Results and discussion

In total, 117 language students completed the survey with 38 participants identifying as Chinese, and 56 as UK. Six participants who identified as Chinese engaged in the diary exercise, which limited comparison opportunities but provided rich insights into these participants' perspectives on SETs. These same participants were also invited to engage in a focus group to further capture in-depth perspectives, though these data were not included in the analysis. Table 1 below provides an overview of responses to data collection tools.

Table 1. Participant nationality groupings and involvement in data collection methods

Nationality groupings	Total	Completed survey	Diary exercise	Focus group
Chinese	38	38	6	6
UK	56	56	0	0
Other	23	23	1 (not analysed)	0

RQ1: To what extent do Chinese and UK students value SET surveys?

Research question 1 was primarily addressed via four questions in the survey. The approach to understanding how students valued SET surveys chiefly involved students' conscious attitudes to SETs, aiming to understand the level of trust students have that completing SETs has some value and that their views should be taken into account in the learning experience. Survey questions 5 and 6a asked students if they had ever completed a SET and if so, what they thought is done with their feedback. The following options could be chosen in question 6a in a multiple response i.e. they could choose more than one.

- a. it is often ignored
- b. it is often used to help improve the course
- c. it is often used to improve teaching
- d. it is often used to market the course i.e. attract new students
- e. it is often used by management to monitor teachers (quality control)

Table 2 shows participant responses to questions 5 and 6a of the survey – the number who chose each and the percentage of that nationality group. Chinese participants had two questions – one on SETs in their own country, one on SETs in the UK. UK participants had one question regarding SETs in the UK.

Table 2. Results of multiple response questions 5 and 6a on what Chinese and UK students believe happens with the results of SETs

Choice: what happens with results	Ignored		To improve course		To improve teaching		To market the course		Management monitoring of teachers	
	n	%	n	%	n	%	n	%	n	%
Chinese in their own country	7	18.4	26	68.4	27	71	5	13.2	13	34.2
Chinese in UK	5	13.2	31	81.6	28	73.7	8	21.1	11	29
UK students	25	44.6	32	57.1	30	53.6	13	23.2	24	42.9

As can be seen in table 2, over two thirds of Chinese participants demonstrated some value in the use of SETs i.e. selecting “to improve teaching/improve the course”, with slightly higher figures for UK versus in their own country. However, UK students showed lower value – percentages in the 50s – while the percentage that selected “ignored” was notably higher (44.6% compared to 13.2% of Chinese students in UK contexts).

A chi-square test of independence (2x2) was performed to examine the relation between nationality and belief that the results are ignored. The relation between these variables was significant, $\chi^2(1, N = 94) = 10.32, p = .001$. UK participants were significantly more likely than Chinese to choose the option that their feedback is often ignored.

A question was then posed in the survey on participants’ views of how honestly they believe they complete SET surveys. The responses are interesting to understand more about student beliefs about SETs and how SETs are used. Survey question 6b examined attitudes when completing SETs, with table 3 (below) showing a comparison between Chinese, UK, and other students (all other nationalities). The overwhelming choice of all nationalities was for “honest feelings” – all were over 60% with Chinese students being the highest at 78.4%.

Table 3. Comparison of responses to question 6b on attitudes to filling in SET surveys by nationality grouping

Choice: attitude to SET completion	I don't usually fill in feedback surveys		I fill surveys in with my honest feelings		I'm kinder than I possibly should be		I'm harsher than I possibly should be		Other	
	n	%	n	%	n	%	n	%	n	%
UK	6	10.7%	35	62.5%	14	25.0%	0	0.0%	1	1.8%
China	2	5.4%	29	78.4%	5	13.5%	1	2.7%	0	0.0%
Other	2	8.7%	16	69.6%	5	21.7%	0	0.0%	0	0.0%

Finally, two survey questions, 7.5 and 7.6, sought to explore whether participants believed their voice should be important in language class design. Question 7.5 “Your teacher should consider YOUR preferences for how language classes should be designed” was aimed at testing a more consumerist orientation, while question 7.6 “Your teacher knows better than you how a language class should be designed” indicates the reverse. The three nationality groupings are compared in table 4 regarding question 7.5 and table 5 regarding question 7.6.

Table 4. Number and percentage of each nationality group choosing each point on disagree – agree scale to the prompt 7.5 “Your teacher should consider YOUR preferences for how language classes should be designed”

Teacher should consider your views	1 strongly disagree		2 disagree		3 neutral		4 agree		5 strongly agree		Mean
	n	%	n	%	n	%	n	%	n	%	
China	0	0	0	0	13	34	19	50	6	16	3.82
UK	0	0	5	9	15	27	20	36	16	26	3.84
Other	0	0	4	17	10	43	7	30	2	9	3.3

Table 5. Number and percentage of each nationality group choosing each point on disagree-agree scale to the prompt 7.6 “Your teacher knows better than you how a language class should be designed” indicates the reverse. The three nationality groupings are compared.”

Teacher knows better	1 strongly disagree		2 disagree		3 neutral		4 agree		5 strongly agree		Mean
	n	%	n	%	n	%	n	%	n	%	
China	0	0	2	5	9	24	16	42	11	29	3.95
UK	2	4	8	14	18	32	20	36	8	14	3.43
Other	2	9	5	22	6	26	4	17	6	26	3.3

A Mann Whitney U test was performed on the Chinese versus UK groups. There was a significant difference between these two national groups in respect of prompt 7.6 (teacher knows better) $U(N \text{ Chinese} = 38, N \text{ UK} = 56) = 765.00, z = -2.414, p = 0.016$. There was no significant difference between the groups for prompt 7.5 (teacher should consider your preferences) – details in table 4. However, for 7.6 – details in table 5 - ordinal logistic regression showed that being of Chinese nationality, contrasted with a UK nationality, increased the odds of scoring higher on the Likert for 7.6 (agreeing with the prompt that your teacher knows better than you how a language class should be designed), by a factor of 1.4, with 95% CI = (0.179, 0.840).

RQ 1: Discussion

Participants' valuation of SETs was approached via prompts that directly addressed their attitudes such as what they are used for and how they approach the surveys they receive (do they complete them; do they complete them honestly, etc). UK participants valued the surveys less, showing a significantly higher belief that SET results are often ignored in comparison with Chinese participants. However, both groups did show a widespread belief that SET results are used to improve teaching and courses. Participants overwhelmingly believe they fill in SETs honestly. There is a lack of research in this area of trust in SETs to help interpret these results.

Of particular interest in the light of the consumerist lens of SETs was the difference in survey responses to the questions about students' role in language class design (questions 7.5 and 7.6), UK participants believe that their views on how to teach language classes are important e.g. UK participants were more in agreement that the teacher should consider their preferences in language class design while Chinese participants had 40% higher odds of saying that the teacher knows better about language lesson design; in other words, Chinese participants have more respect for or trust in the knowledge of the teacher, or perhaps just less trust in their own knowledge; Chinese participants appear to defer to the teacher as best placed to make the judgement on lesson delivery. This is perhaps in line with a difference in self-evaluation that has been observed between East and West. This self-evaluation psychological construct relates to how highly someone rates themselves e.g. in abilities, with UK based participants apparently holding higher self-evaluation than Chinese participants, (Ross & Wang, 2010) i.e. Chinese participants may have a lower evaluation of their own ability to assess quality. Other theories which might inform this difference concern the role of and respect for authority in China, linked to the teacher and their knowledge; if well-presented, students will expect to reproduce or at least closely resemble what experts deliver i.e. authorised knowledge – in basic terms, the teacher is the expert (Pratt et al., 1999, p. 252).

Free-text comments in the survey provide some interesting colour to the above points in respect of UK/Western attitudes. For example, one UK participant said, "teach us how to speak a language, don't teach us frankly petty and unimportant grammar points".

The above participant statement indicates a judgement on the importance or otherwise of points of grammar suggesting a belief that student views on the language are, at the very least, important for the teacher to consider and could indicate a consumerist orientation. The consumerist lens on SETs has arguably become the norm in UK higher education i.e. students' satisfaction is of central importance (Bunce et al., 2017; Lomas, 2007). However, the underlying assumption in SETs that students know what quality is (Stroebe 2020) is problematic, with quality being highly relative – e.g. it is "in the eye of the beholder" (Vroeijenstijn, 1995, cited in Watty, 2006, p. 292). As we saw in sections 1 and 2, satisfaction is influenced by service-user aims/wants (not necessarily needs) and wants in language learning may be based on misplaced learner beliefs and indeed these may actually damage their learning and progress (Horwitz, 1988). The results here show that in this context, UK participants were significantly less likely to defer to teacher knowledge/expertise i.e. they would appear to view their role in negotiating classroom choices more stridently and suffer a negative disconfirmation under the satisfaction framework of service feedback i.e. not meeting expectations (Sultan & Yin Wong, 2012).

A problem worthy of further study, in addition to the language learning beliefs problem above, is the possibility that a student's consumer self-view is linked to academic performance – students with a high consumer self-view are less likely to engage in the intellectual pursuit of knowledge and consequently perform worse (Bunce et al., 2017). Another consequence of the consumerist framework, it is claimed, is worse teaching with teachers primarily concerned with improving student satisfaction; satisfaction being improved, so the claims go, by giving students less work and generally making courses easier (Stroebe, 2020).

RQ 2: Do Chinese and UK students focus on different attributions (causes) and different loci of effect when presented with certain prompts used in SET surveys – satisfied; attentive; good quality; engaged?

Survey questions 7.1 to 7.4, and 8-12 were designed to provide insights into participant attributions for positive and negative classroom experiences, with diaries also used to understand more about attributions by asking participants to complete entries starting “I was/wasn't satisfied with the quality of the lesson and teaching because...”. Of interest was whether participants show a bias towards the teacher being responsible for positive classroom experiences and bearing responsibility/blame for negative classroom experiences, or a bias towards their own actions and orientation i.e. do they tend towards external locus (probably uncontrollable, possibly stable) of attribution in these situations.

Informing the design of survey questions and the analysis of diary entries was the analytical framework shown in table 6. It is built on three dimensions to attribution: locus, stability and controllability. This framework was developed by Graham (1991) from earlier work by Weiner (Graham, 1991, p. 7), with the version shown here also informed by Chaparro et al. (2023).

Table 6. Framework to analyse survey responses to questions 7.1 – 7.4, 8 – 12, and student diaries

	Internal locus		External locus	
	Stable -> Unstable		Stable -> Unstable	
Controllable	Student willingness to participate and typical level of effort	Student participation and effort now; Energy levels (due to diet and sleep for example)	Teacher's methods e.g. active learning approach; didactic approach	Fellow-students; teaching space
Energy levels (due to diet and sleep for example)	Intelligence; character	Mood; energy level (default or biological)	Task difficulty Teacher's character	Teacher's mood; luck; weather

Locus of attribution may be internal or external. In the context of student feedback, this would concern whether the attribution/cause of any positive or negative experiences rests with the participants themselves (internal) or rests with the teacher, environment, and so on (external). Within each of the two loci we can move to the dimension of stability i.e. is the cause something which changes (unstable) or is relatively fixed (stable); in other words, this concerns how dynamic the relationship is between an outcome and its cause. Thus, an internal and stable locus is a student's intelligence whereas their willingness to participate is more likely to change. There is a third dimension, which is the level of controllability (in education this means from the student's perspective) i.e. loci can be viewed as controllable or uncontrollable in the sense that students can influence/exert control over the causes of their experiences. Effort is something a student should see as controllable, whereas intelligence, their character, and the teacher's character are probably not (Graham, 1991).

If Chinese (Eastern) nationals exhibit the type of different self-construal biases compared to Western nationals as introduced in the section on cultural biases above (e.g. Kitayama et al., 2009; Lewis et. al, 2008) we are less likely to see uncontrollable external causes linked to negative experiences. Chinese participants should be more likely than UK students to consider context and less likely to demonstrate self-serving bias (blaming or attributing failures to others) (Choi et al., 1999; Mezulis et al., 2004).

Survey prompts 7.1 – 7.4 examined this via a 5-point disagree-agree scale (1 = strongly disagree, 5 = strongly agree). They were as follows:

- 7.1 If you enjoyed lessons, it was due to your teacher
- 7.2 If you did not enjoy lessons, it was due to your teacher
- 7.3 If you enjoyed lessons, it was due to you preparing well or being motivated
- 7.4 If you did not enjoy lessons, it was due to you not preparing well or not being motivated

Data were analysed to establish if there was a nationality association with prompts: 7.1 and 7.2 indicate an external locus for causes/attributions; 7.3 and 7.4 focus on the participants themselves and so suggest an internal locus for causes/attributions.

Table 7. Comparison of means on scale disagree-agree that students' enjoyment and non-enjoyment of lessons was due to the teacher.

Survey question means	Qu 7.1 mean	Qu 7.2 mean	Qu 7.3 mean	Qu 7.4 mean
China	3.76	2.79	4	3.08
UK	3.98	3.3	3.79	3.25
Other	4.3	3.96	3.7	3.3

Table 7 shows comparison of participants' mean agreement scores, with UK participants showing higher agreement than Chinese participants with prompt 7.1 (positive experience,

external cause), and 7.2 (negative experience, external cause) and 7.4 (negative experience, internal cause). A Mann Whitney U test was performed on the Chinese versus UK groups to compare responses to each prompt. A significant difference was found in only one comparison, between Chinese and UK responses to question 7.2 (that attribution lies with the teacher when the outcome is lack of enjoyment): $U(N \text{ Chinese} = 38, N \text{ UK} = 56) = 1323.50, z = 2.066, p = 0.039$ i.e. UK participants are much more likely to blame the teacher for a negative experience. It is interesting to observe that the other nationality category (mostly made up of Western nationalities) was even more marked in its agreement with 7.2 than UK participants.

Chinese participant diaries were analysed to shed further insight on the type of experiences they focus on – positive or negative – and the attribution for these. Template analysis method was used with a priori code (Symon et. al, 2012). The template was constructed through a lens of self-serving/hedonistic bias using the framework in table 6. Language was identified that connected to attributions for the prompts, such as internal (personal) and external and also on effect on the individual or on others. Table 8 presents the key themes arising when viewing student texts through this template. Participants did not always follow the instructions e.g. the same prompt was used more than twice by some (hence there were more entries for some prompts compared to others). Results show a clear predominant focus on positive experiences, attributed to the teacher.

Table 8. Summary of diary analysis based on locus of effects and locus of attributions

Key word/ prompt	Number of diary entries using this prompt	Effect – positive or negative and focus of memory e.g. focus on “me” focus on “us”	Explanation/ Attribution	Students (number of students’ entries fitting into this theme)
1. Satisfied with quality	15	Positive experience with effect on me or us	Teacher action	10
			Holistic or group action	7
			Personal student action	1
		References to fun/enjoyment		1
		Negative experience with effect on me or us	Teacher action	3
			Holistic or shared responsibility	2
		References to wasting time or similar		2
2. Engaged	10	Positive experience effect on me or us	Teacher action	9
			Holistic or group action	2
			Personal student action	1
		Negative experience with effect on me or us	Personal student action	1

3. Teacher explained things well	13	Positive experience effect on me or us	Teacher action	13
			Holistic or group action	6
			Personal student action	1
4. Attentive	9	Positive experience effect on me or us	Teacher action	6
			Holistic or group action	3
		Negative experience effect on me or us	Teacher action	1

To provide additional insight into external and internal locus attribution and to whether participants focus on positive or negative classroom experience, the survey contained an image association task. Four classroom photographs (see figure 1 below) were used for questions 8-11:

Question 8 Which picture would you choose for a student who commented on the quality of the course?

Question 9 Which picture would you choose for a student who commented on being engaged, or not being engaged?

Question 10 Which picture would you choose for a student who commented on being satisfied or unsatisfied?

Question 11 Which picture would you choose for a student who commented on being attentive or not attentive in class?

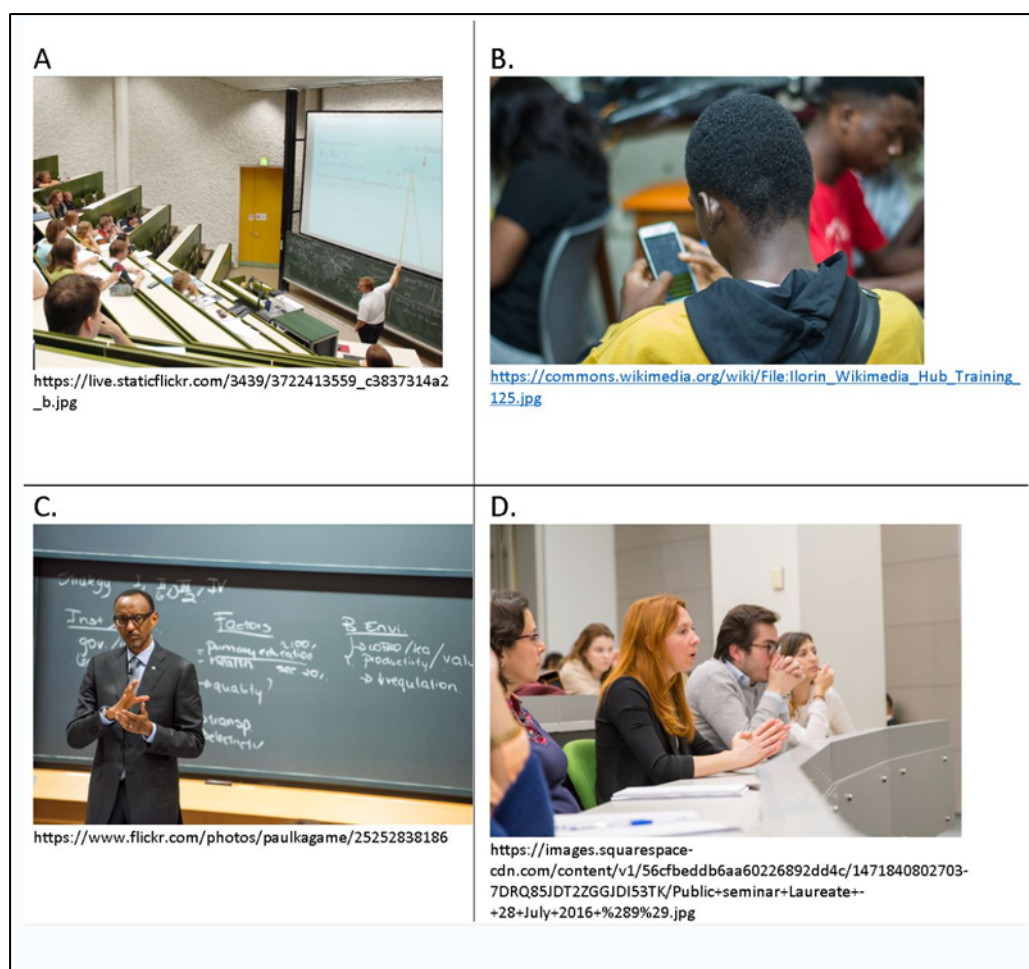


Figure 1. Images used for picture association prompts

It was hypothesised based on what we know about the differences between East and West and attribution (e.g. Choi et al., 1999; Lewis et al., 2008; Mezulis et al. 2004), that UK participants might be more likely than Chinese to choose a student-focussed picture for a positive experience and teacher-focussed picture for a negative experience i.e. linking positive experiences to internal causes. However, in questions 9, 10, and 11, participants could choose to take the positive or negative from the question prompt e.g. question 10 said “Which picture would you choose for a student who commented on being engaged, or not being engaged?” A flaw with the choice of pictures, not picked up during piloting, was that the two teacher focussed pictures did not clearly show which was a positive or negative experience. Due to this weakness and also low counts for some picture choices, areas chosen to focus on were:

- any association between national group’s (China and UK) preference for choosing a student-focussed picture (B or D) versus a teacher-focussed picture (A or C)
- any association between national group’s (China and UK) preference for choosing a negative (picture B) or positive (picture D) student experience.

Table 9 shows the results of the picture association survey questions 8 - 11.

Table 9. Number and percentage of Chinese and UK participants choosing one of four pictures (A-D) for key-word prompts “quality”, “engaged”, “satisfied”, “attentive”

	Chinese pic A	UK pic A	Chinese pic B	UK pic B	Chinese pic C	UK pic C	Chinese pic D	UK pic D
8. quality	12, 31.6%	n16, 29.1%	n4, 10.5%	n2, 3.6%	n5, 13.2%	n13, 23.6%	n17, 44.7%	n24, 43.6%
9. engaged	n7, 18.4%	n8,14.5%	n9, 23.6%	n28, 50.9%	n5,13.2%	n0, 0%	n17, 44.7%	n19, 34.5%
10. satisfied	n10, 26.3%	n13, 23.6%	n9, 23.6%	n17, 30.9%	n4, 10.5%	n6, 10.9%	n15, 39.5%	n19, 34.5%
11. attentive	n7, 18.9%	n5, 9.4%	n18, 48.6%	n28 52.8%	n1, 2.7%	n4, 7.5%	n11, 29.7%	n16, 30.2%
Total selections of this picture	n36, 23.8%	n42, 19.3%	n40, 26.5%	n75, 34.4%	n15, 9.9%	n23, 10.6%	n60, 39.7%	n78, 36%

N Chinese = 38 except picture 11 where N Chinese = 37; N UK = 55 except picture 11 where N UK = 53

In terms of a) – any association between national groups and choice of effect i.e. choosing a picture centred on a student rather than something else (class/classroom or teacher) – table 9 shows differences but no significance in differences between participants linking prompts to particular picture categories (B and D student-focussed; A and C teacher-focussed).

Prompt 8 – a student commented on the quality of the course - was almost equally associated with a student experience (pictures B and D) and a teacher focused image (A and C). Overall, the major choice was a positive student experience (picture D) but there was no significance from a Chi square test. Also, for prompt 9 - a student being engaged or not engaged - showed no significant difference in association by Chinese or UK participants, though it was overwhelmingly associated by all participants with a student experience (pictures B and D) rather than with context or the teacher (pictures A and C). Prompt 10 was similarly more associated with a student experience or focus. Prompt 11 was perhaps interesting in that most participants seemingly took an “inattentive” rather than “attentive” interpretation if we take the student’s use of a mobile phone as their not attending to the teaching.

This leads us to consider b – any association between national group’s (China and UK) preference for choosing a negative (picture B) or positive (picture D) student experience. Table 9 shows a marked difference in the Chinese group focussing on a positive student experience (picture D) compared to a negative experience (picture B): around 66% of Chinese participants selected a student-centred picture (B and D), with almost 40% choosing a positive student experience and 26.5% negative. This contrasts with the UK group where the percentages choosing each are around 35% choosing positive and negative student experience pictures. Contrasting the 26.5% negative for Chinese with 35% negative for UK, a Chi square test of independence (2x2) was performed to examine the difference; the results did not show significance $\chi^2(1, N=94) = 4.152, p = 0.074$. The problem here may be the number of counts below 5 to satisfy Chi square requirements.

However, there is the suggestion from the diary exercise that there may be a tendency for Chinese participants to focus more on positive outcomes.

To provide further depth in understanding what the prompts activate in participants' minds, Chinese participants were compared with UK participants in their choice of words to associate with the key prompts satisfied, attentive, good quality, engaged. Western students, according to theory, are more likely to choose a taxonomic relationship in word association, suggesting they focus more on isolated objects and attributes compared to Chinese thinking more holistically incorporating relationships to context and environment (Estes et al., 2011). Any significant differences found here might suggest it being worthwhile in future to explore what students think about in terms of effect when they are asked to think about locus of attribution.

To assist analysis of participant word associations (satisfied, attentive, good quality, engaged), the framework below was used, adapted from Saalbach and Imai's work (2007), based on Ji et al. (2004); it uses taxonomic and theme association types. Table 10 shows the percentage of UK and Chinese participants choosing a word that fits each category, together with the raw figures. The taxonomic column for satisfied, attentive, and engaged was used for words linked to emotions or internal actions i.e. actions of the student; the theme column for these prompts was used for words linked to context or anything external to the student themselves. An example is provided to illustrate the logic of the system.

Table 10. Percentage of UK and Chinese participants choosing a word fitting under each category. Percentages have been rounded to the nearest whole number (up or down).

Prompt	Nationality	Taxonomic	Holistic - Theme	Does not fit, not clear	Total count (N)
Example					
Satisfied	x	Happy; listening; contented; enjoy; unsatisfied; unhappy	Party; nice music; communication; space		
Satisfied	UK	n49, 92%	n3, 6%	n1, 2%	53
	Chinese	n18, 69%	n6, 23%	n2, 8%	26
Attentive	UK	n45, 85%	n8, 15%	n0, 0%	53
	Chinese	n9, 31%	n19, 66%	n1, 3%	29
Good quality	UK	n49, 92%	n3, 6%	n1, 2%	53
	Chinese	n14, 50%	n13, 46%	n1, 4%	28
Engaged	UK	n46, 87%	n7, 13%	n0, 0%	53
	Chinese	n13, 50%	n13, 50%	n0, 0%	26

Some examples of words supplied by participants in the survey are as follows in table 11.

Table 11. Some survey responses providing examples for each prompt

Taxonomic for satisfied	Holistic for satisfied
Happy, unsatisfied	Learning, course content
Taxonomic for attentive	Holistic for attentive
Engaged, boring	Interaction, tutor
Taxonomic for good quality	Holistic for good quality
Motivated, focused	Thought providing
Taxonomic for engaged	Taxonomic for engaged
Interested, engaged	Homework; share opinion

The word association task data in table 10 shows a clear tendency for Chinese participants to link the prompts to thematic or holistic concepts, while UK participants are much more likely to connect the prompts to a taxonomy which in practice meant other words associated with the feelings or actions of students (in the case of satisfied, attentive, and engaged).

RQ 2: Discussion

Many of the data linked to research question 2 point to a cultural difference in how Chinese and UK students interpret SET prompts: effects/outcomes and attributions for these. According to Ross & Wang (2010), culture influences what we perceive or focus on e.g. in a physical scene, what we remember and how we use memory. Under the concept of autobiographical memory, Eastern cultures, unlike Western, do not put themselves at the centre of events (Ross & Wang, 2010). Chinese participants made many external (including group) attributions in these “social events” (e.g. see Morris & Peng, 1994) for positive experiences, in contrast to hedonistic bias expectations applicable to Western students. Taking into account the picture association task in the survey, two national differences between UK and Chinese participants were observed: Chinese participants were more likely to focus on a positive student experience than a negative one and when given the choice of an engaged or unengaged student, the UK participants were more likely to choose the unengaged student i.e. the negative experience when prompted with “engaged”.

Based on directly questioning on attribution (survey questions 7.1 to 7.4), in line with the general East-West difference in bias – particularly hedonistic bias (Maddux & Yuki, 2006; Mezulis et al., 2004), UK participants were significantly more likely to view the cause of negative experiences as being the teacher’s actions (an external locus of attribution) compared to Chinese participants. This complements research in Western service contexts, as marketing literature has conceptualised education (e.g. Athiyaman, 1997), where service failures are more likely to be attributed to external causes and successes to internal causes (Athiyaman, 1997).

The diary activity supported these observations with Chinese participants tending to talk about positive experiences because of what the teacher did, not what they did themselves, and holistic perspectives were also common e.g. talking about a good “atmosphere” or describing group activities. Triangulated with the word-association task, there is a clear suggestion that Chinese participants were less likely, with the exception of the prompt “satisfied”, to focus on the individual (themselves/the student). This is consistent with other research; Oishi (2002, p. 1398) states “East Asians are accustomed to thinking about both sides of a phenomenon and therefore they are likely to view both negative and positive aspects of any event as valid and consciously take into account both positive and negative information when making global judgments.” (Oishi, 2002, p. 1399). Coming back to the prompt “satisfied”, although it is indicative of an individualistic perspective for Chinese and UK students, the picture association task suggests this is more likely to be associated with a positive effect/outcome by Chinese participants. By contrast, taking into consideration the triangulation of survey questions 7.1-7.4, the picture association, and the word association tasks, UK participants were more likely to focus on the individual (thinking about themselves) for all prompts i.e. less holistically and less evidently skewed towards positive experiences. It is generally accepted that there are limitations in word association such as in the subjectivity in how responses are interpreted e.g. how word responses to cues are grouped or labelled, hence the advisability of more than one projective method – e.g. in this case, the use of pictures (Donoghue, 2000).

It is interesting to consider again fundamental attribution error at this point when looking at the personal (the student and the teacher) in situations. Not only does the data suggest UK participants were more likely to think about the personal in terms of effect/outcome, they may be more likely to focus on the personal in attribution i.e. prompt 7.2 showed significance in difference between UK and Chinese with the former more likely to place locus on the teacher; fundamental bias points towards overstating personal characteristics in others for situations that affect us (Ross & Nisbett, 2011). Two free-text remarks from Western participants, one from an American and one from the UK provide some texture here although should not be overstated – both respondents indicated they believed negative/constructive SET feedback was ignored due to the personality of the teacher: “negative feedback is met with defensiveness” and “a lot of lecturers especially have big egos”. Worthy of further study in SET contexts is whether there is the risk that with Western students, their reaction when faced with an experience they are not happy with is to attribute the cause to the teacher not liking them or some other aspect of the teacher’s character.

In summary regarding question 2, we should not assume the prompts satisfied, attentive, engaged, and quality have parity of meaning between East and West, though it is difficult to tell if particular terms are preferable for SETs. All appear to encourage a focus on the effect on the student (not classmates or the class atmosphere etc), more so with UK students. Also, there appear to be differences in participants’ tendency to focus on a positive or negative and who/what to attribute this to. With UK participants appearing to show self-serving or hedonistic bias in the SET context i.e. to blame negative experiences on others, there are important implications for language learning. Personal autonomy is considered important for all learning: “enhanced motivation is conditional on learners taking responsibility for their own learning [...] and perceiving that their learning successes and failures are to be attributed to their own efforts and strategies rather than

to factors outside their control” (Dickinson, 1995, cited in Dörnyei & Csizér, 1998, p. 217). The hedonistic tendency may be further exacerbated when seeking feedback that is obviously consumer orientated, using questions such as were you satisfied, which may promote a consumer identity, while reducing personal responsibility. In addition, there would be risks with implementing changes as a response to feedback if the feedback is from students with a higher consumer orientation (Bunce et al., 2017).

Conclusion and implications for practice

This study has attempted to develop new understanding of student feedback on language teaching in higher education with a particular focus on end-of-course survey methods. In particular, the project has examined whether evaluation surveys and their questions/prompts, particularly the key words satisfied and engaged, produce useful information and have a positive impact as assessed through two conditions. Firstly, do students explicitly value them? Secondly, for classes dominated by Chinese students or UK students, are SET results equally valid; if, for example, there are large numbers of Chinese students present on courses, do they exhibit different tendencies in their engagement with feedback that undermine attempts to compare classes or courses with a different nationality mix?

In language teaching contexts this research demonstrates sufficient grounds to approach with caution SET data from predominantly UK students as being comparable to SET data from classes or courses with large numbers of Chinese. Chinese and UK participants in this study showed differences in attribution bias and autobiographical memory, with UK participants more likely to focus on how teaching affects them personally and to attribute a negative experience to the actions or character of the teacher. Indeed, UK participants may be more likely to focus on negatives in the first place.

I would argue that key prompts often used in SETs such as “engaged” and “satisfied” encourage a consumer mindset and are inconsistent with learning theories that value autonomy (e.g. Dörnyei & Ryan 2015). The removal of the NSS prompt “Overall, I am satisfied with the quality of the course” for England (though not for Scotland) has this as part of the rationale (OfS, 2022), though QAA have expressed opposition to this move (QAA, 2022). If SETs aim, among other things, to empower students, they should also encourage students to view themselves as having control, especially to view outcomes as linked to effort (pushing students to place a more dynamic internal controllable attribution to outcomes).

The case study dimensions of the project limit transferability of the results but clearly raise important questions about the assumptions that go into feedback prompts and questions about use of survey approaches, especially at the end of courses. Pragmatically, the following suggestions are made in consequence of these research findings on how to maximise student trust in SETs and improve their value.

When designing and using survey SETs, we should think about:

- a. how prompts are framed to recognise cultural biases and difference in response and engagement
- b. including prompts that promote student self-reflection as to their agency/role in teaching quality

- c. whether we wish students to consider impacts not only on themselves but on others
- d. how we might interpret or triangulate information from surveys e.g. via different prompts and different modes (i.e. approaches other than surveys to increase engagement and allow for more qualitative information)
- e. teachers providing more explicitness about teaching and learning approaches, particularly in language classrooms as the communicative approach links to a holistic impact that students need to consider beyond individual impact and preferences.

As Sharpe (2019, p. 38) says “Surveys appear to be so well ingrained into the management of teaching and learning that it is unlikely they will be removed”. However, understanding more about the data and understanding how to improve it is clearly vital if we are to empower both students and teachers.

SETs are intended to empower students, yet this study is in line with others (e.g. Denson et al., 2010) that there are questions as to the effectiveness of survey SETs and their prompts for students from different cultures where this is not considered in the design. Further research in this area with larger sample sizes and across a variety of subject contexts would seem important. Also, it would be interesting to examine whether biases may be more or less evident from personal emotion prompts like satisfied and engaged when employed in feedback modes other than surveys (and diaries) e.g. in one-to-one conversations with students.

Declarations

The author discloses that they have no actual or perceived conflicts of interest. The author discloses that they have not received any funding for this manuscript beyond resourcing for academic time at their respective university. The author has not used artificial intelligence in the ideation, design, or write-up of this research.

References

- Allison, N. G. (2020). Students' attention in class: Patterns, perceptions of cause and a tool for measuring classroom quality of life. *Journal of Perspectives in Applied Academic Practice*, 8(2), 58–71. <https://doi.org/10.14297/jpaap.v8i2.427>
- Angell, R. J., Heffernan, T. W., & Megicks, P. (2008). Service quality in postgraduate education. *Quality Assurance in Education*, 16(3), 236–254. <https://doi.org/10.1108/09684880810886259>
- Angelo, T. A., & Cross, K. P. (1993). *Classroom assessment techniques: A handbook for college teachers* (2nd ed.). Jossey-Bass.
- Angelo, T. A., & Zakrajsek, T. D. (2024). *Classroom assessment techniques: Formative feedback tools for college and university teachers*. John Wiley & Sons.
- Arthur, J., Waring, M., Coe, R., & Hedges, L. V. (2012). *Research methods and methodologies in education*. SAGE.
- Athiyaman, A. (1997). Linking student satisfaction and service quality perceptions: The case of university education. *European Journal of Marketing*, 31(7), 528–540. <https://doi.org/10.1108/03090569710176655>

Brown, A. (2009). Students' and teachers' perceptions of effective foreign language teaching: A comparison of ideals. *The Modern Language Journal*, 93(1), 46–60. <https://doi.org/10.1111/j.1540-4781.2009.00827.x>

Bunce, L., Baird, A., & Jones, S. E. (2017). The student-as-consumer approach in higher education and its effects on academic performance. *Studies in Higher Education*, 42(11), 1958–1978. <https://doi.org/10.1080/03075079.2015.1127908>

Chalmers, D., & Hunt, L. (2016). Evaluation of teaching. *Higher Education Research and Development*, 3, 25–55. www.hersa.org.au/herdsa-review-higher-education-vol-3/25-55

Chaparro, R., Reaves, M., Jagger, C. B., Bunch, J. C. (2023, April 24) Attribution Theory: How Is It Used? Askifas. <https://edis.ifas.ufl.edu/publication/WC162>

Choi, I., Nisbett, R. E., & Norenzayan, A. (1999). Causal attribution across cultures: Variation and universality. *Psychological bulletin*, 125(1), 47. <https://doi.org/10.1037/0033-2909.125.1.47>

Cibangu, S. K., & Hepworth, M. (2016). The uses of phenomenology and phenomenography: A critical review. *Library & Information Science Research*, 38(2), 148–160. <https://doi.org/10.1016/j.lisr.2016.05.001>

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn?: A meta-analysis and review of the literature. *Journal of Marketing Education*, 31(1), 16–30. <https://doi.org/10.1177/0273475308324086>

Danioni, F., Coen, S., Rosnati, R., & Barni, D. (2020). The relationship between direct and indirect measures of values: Is social desirability a significant moderator?. *European Review of Applied Psychology*, 70(3), 100524. <https://doi.org/10.1016/j.erap.2020.100524>

Davidson, A. S. (2013). Phenomenological approaches in psychology and health sciences. *Qualitative Research in Psychology*, 10(3), 318–339. <https://doi.org/10.1080/14780887.2011.608466>

Denson, N., Loveday, T., & Dalton, H. (2010). Student evaluation of courses: What predicts satisfaction? *Higher Education Research & Development*, 29(4), 339–356. <https://doi.org/10.1080/07294360903394466>

Diamond, A., Evans, J., & Sheen, J. (2015). UK review of information about higher education: Information mapping study. <https://www.voced.edu.au/content/ngv:72978>

Donoghue, S. (2000). Projective techniques in consumer research. *Journal of Consumer Sciences*, 28(1). <https://doi.org/10.4314/jfecs.v28i1.52784>

Dörnyei, Z., & Csizér, K. (1998). Ten commandments for motivating language learners: Results of an empirical study. *Language Teaching Research*, 2(3), 203–229. <https://doi.org/10.1177/136216889800200303>

Dörnyei, Z., & Ryan, S. (2015). *The psychology of the language learner revisited*. Routledge. <https://doi.org/10.4324/9781315779553>

Drożdżiał-Szelest, K., & Pawlak, M. (2013). *Psycholinguistic and sociolinguistic perspectives on second language learning and teaching: Studies in honor of Waldemar Marton*. Springer. <https://doi.org/10.1007/978-3-642-23547-4>

Duisembekova, Z., & Özmen, K. S. (2020). Analyzing language learning beliefs of English student teachers: A cross-cultural study across Turkic republics. *Bilig*, 94, 51–73.

Ellis, R. (2012). *Language teaching research and language pedagogy*. Wiley-Blackwell.

<https://doi.org/10.1002/9781118271643>

Estes, Z., Golonka, S., & Jones, L. L. (2011). Thematic thinking: the apprehension and consequences of thematic relations. *The Psychology of Learning and Motivation*, 54, 249–294.

<https://doi.org/10.1016/B978-0-12-385527-5.00008-5>

Feistauer, D., & Richter, T. (2017). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, 42(8), 1263–1279.

<https://doi.org/10.1080/02602938.2016.1261083>

Ghedin, E., & Aquario, D. (2008). Moving towards multidimensional evaluation of teaching in higher education: A study across four faculties. *Higher Education*, 56(5), 583–597.

<https://doi.org/10.1007/s10734-008-9112-x>

Graham, S. (1991). A review of attribution theory in achievement contexts. *Educational Psychology Review*, 3(1), 5–39. <https://doi.org/10.1007/BF01323661>

Gunn, A. (2018). Metrics and methodologies for measuring teaching quality in higher education: Developing the Teaching Excellence Framework (TEF). *Educational Review*, 70(2), 129–148.

<https://doi.org/10.1080/00131911.2017.1410106>

Harvey, L., & Green, D. (1993). Defining quality. *Assessment and Evaluation in Higher Education*, 18(1), 9–34. <https://doi.org/10.1080/0260293930180102>

Heffernan, T. (2022). Sexism, racism, prejudice, and bias: A literature review and synthesis of research surrounding student evaluations of courses and teaching. *Assessment & Evaluation in Higher Education*, 47(1), 144–154. <https://doi.org/10.1080/02602938.2021.1888075>

Hoffman, C., Lau, I., & Johnson, D. R. (1986). The linguistic relativity of person cognition: An English–Chinese comparison. *Journal of Personality and Social Psychology*, 51(6), 1097.

<https://doi.org/10.1037/0022-3514.51.6.1097>

Horwitz, E. K. (1988). The beliefs about language learning of beginning university foreign language students. *The Modern Language Journal*, 72(3), 283–294. <https://doi.org/10.1111/j.1540-4781.1988.tb04190.x>

ICEF Monitor. (2022, March 31). Measuring the economic impact of foreign students in the UK and the country's competitive position in international recruitment. *ICEF Monitor*.

<https://monitor.icef.com/2021/09/measuring-the-economic-impact-of-foreign-students-in-the-uk-and-the-countrys-competitive-position-in-international-recruitment/>

Iida, M., Shrout, P. E., Laurenceau, J.-P., & Bolger, N. (2012). Using diary methods in psychological research. *APA Handbook of Research Methods in Psychology, Vol 1: Foundations, Planning, Measures, and Psychometrics* (pp. 277–305). American Psychological Association.

<https://doi.org/10.1037/13619-016>

Ji, L.-J., Zhang, Z., & Nisbett, R. E. (2004). Is It culture or is it language? Examination of language effects in cross-cultural research on categorization. *Journal of Personality and Social Psychology*, 87(1), 57–65. <https://doi.org/10.1037/0022-3514.87.1.57>

Jung, C. G. (1910). The association method. *The American Journal of Psychology*, 21(2), 219–269.

<https://doi.org/10.2307/1413002>

- Kitayama, S., Park, H., Sevincer, A. T., Karasawa, M., & Uskul, A. K. (2009). A cultural task analysis of implicit independence: comparing North America, Western Europe, and East Asia. *Journal of Personality and Social Psychology*, 97(2), 236–255. <https://doi.org/10.1037/a0015999>
- Langan, A. M., & Harris, W. E. (2019). National student survey metrics: Where is the room for improvement? *Higher Education*, 78(6), 1075–1089. <https://doi.org/10.1007/s10734-019-00389-1>
- Lewis, R. S., Goto, S. G., & Kong, L. L. (2008). Culture and context: East Asian American and European American differences in P3 event-related potentials and self-construal. *Personality and Social Psychology Bulletin*, 34(5), 623–634. <https://doi.org/10.1177/0146167207313731>
- Li, C., Lu, J., Bernstein, B., & Bang, N. M. (2018). Counseling self-efficacy of international counseling students in the US: The impact of foreign language anxiety and acculturation. *International Journal for the Advancement of Counselling*, 40(3), 267–278. <https://doi.org/10.1007/s10447-018-9325-3>
- Loewen, S., Li, S., Fei, F., Thompson, A., Nakatsukasa, K., Ahn, S., & Chen, X. (2009). Second language learners' beliefs about grammar instruction and error correction. *The Modern Language Journal*, 93(1), 91–104. <https://doi.org/10.1111/j.1540-4781.2009.00830.x>
- Lomas, L. (2007). Are students customers? Perceptions of academic staff. *Quality in Higher Education*, 13(1), 31–44. <https://doi.org/10.1080/13538320701272714>
- Luo, Y., Xie, M., & Lian, Z. (2019). Emotional engagement and student satisfaction: a study of Chinese College students based on a nationally representative sample. *The Asia-Pacific Education Researcher*, 28(4), 283–292. <https://doi.org/10.1007/s40299-019-00437-5>
- Maddux, W. W., & Yuki, M. (2006). The “ripple effect”: Cultural differences in perceptions of the consequences of events. *Personality and Social Psychology Bulletin*, 32(5), 669–683. <https://doi.org/10.1177/0146167205283840>
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76(5), 707–754. <https://doi.org/10.1037/0022-0663.76.5.707>
- Mavondo, F. T., Tsarenko, Y., & Gabbott, M. (2004). International and local student satisfaction: Resources and capabilities perspective. *Journal of Marketing for Higher Education*, 14(1), 41–60. https://doi.org/10.1300/J050v14n01_03
- Mayring, P. (2014, August). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. SSOAR Open Access Repository. <https://www.ssoar.info/ssoar/handle/document/39517>
- Mesías, F. J., & Escribano, M. (2018). Chapter 4—Projective Techniques. In G. Ares & P. Varela (Eds.), *Methods in Consumer Research, Volume 1* (pp. 79–102). Woodhead Publishing. <https://doi.org/10.1016/B978-0-08-102089-0.00004-2>
- Mezulis, A. H., Abramson, L. Y., Hyde, J. S., & Hankin, B. L. (2004). Is there a universal positivity bias in attributions? A meta-analytic review of individual, developmental, and cultural differences in the self-serving attributional bias. *Psychological Bulletin*, 130(5), 711. <https://doi.org/10.1037/0033-2909.130.5.711>
- Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology*, 67(6), 949–971. <https://doi.org/10.1037/0022-3514.67.6.949>

Office for Students (OfS). (2022). Consultation on changes to the National Student Survey.

<https://www.officeforstudents.org.uk/media/93f31b21-c1ee-469d-adeb-87df264d2034/consultation-on-changes-to-national-student-survey.pdf>

Office for Students (OfS). (2023a). National Student Survey – NSS.

<https://www.officeforstudents.org.uk/data-and-analysis/national-student-survey-data/nss-data-archive/nss-2023-results/>

Office for Students (OfS). (2023b). The TEF - a guide for students.

<https://www.officeforstudents.org.uk/for-students/teaching-quality-and-tef/the-tef-a-guide-for-students/>

Oishi, S. (2002). The experiencing and remembering of well-being: A cross-cultural analysis.

Personality and Social Psychology Bulletin, 28(10), 1398–1406.

<https://doi.org/10.1177/014616702236871>

Pineda, P., & Seidenschnur, T. (2021). The metrification of teaching: Student evaluation of teaching in the United States, Germany and Colombia. *Comparative Education*, 57(3), 377–397.

<https://doi.org/10.1080/03050068.2021.1924447>

Playfoot, D., Balint, T., Pandya, V., Parkes, A., Peters, M., & Richards, S. (2018). Are word association responses really the first words that come to mind? *Applied Linguistics*, 39(5), 607–624.

<https://doi.org/10.1093/applin/amw015>

Prakash, G. (2018). Quality in higher education institutions: insights from the literature. *TQM Journal*, 30(6), 732–748.

<https://doi.org/10.1108/TQM-04-2017-0043>

Pratt, D. D., Kelly, M., & Wong, W. S. (1999). Chinese conceptions of ‘effective teaching’ in Hong Kong: Towards culturally sensitive evaluation of teaching. *International Journal of Lifelong Education*, 18(4), 241–258.

<https://doi.org/10.1080/026013799293739a>

Quality Assurance Agency (QAA). (2022, August 22). QAA calls for the NSS to keep asking students in England about satisfaction with the quality of their course. QAA.

<https://www.qaa.ac.uk/news-events/news/qaa-calls-for-the-nss-to-keep-asking-students-in-england-about-satisfaction-with-quality-of-course>

Ross, L., & Nisbett, R. E. (2011). *The person and the situation: Perspectives of social psychology*.

Pinter & Martin Publishers.

Ross, M., & Wang, Q. (2010). Why we remember and what we remember: culture and autobiographical memory. *Perspectives on Psychological Science*, 5(4), 401–409.

<https://doi.org/10.1177/1745691610375555>

Saalbach, H., & Imai, M. (2007). Scope of linguistic influence: Does a classifier system alter object concepts? *Journal of Experimental Psychology: General*, 136(3), 485.

<https://doi.org/10.1037/0096-3445.136.3.485>

Sale, J. E., Lohfeld, L. H., & Brazil, K. (2002). Revisiting the quantitative-qualitative debate:

Implications for mixed-methods research. *Quality and quantity*, 36, 43–53.

<https://doi.org/10.1023/A:1014301607592>

Sharpe, R. (2019). Evaluating the student experience: a critical review of the use of surveys to enhance the student experience. In K. Trimmer, T. Newman, & F. F. Padró (Eds.), *Ensuring Quality in Professional Education Volume II: Engineering Pedagogy and International Knowledge Structures* (pp. 29–45). Springer International Publishing. https://doi.org/10.1007/978-3-030-01084-3_2

Shepperd, J., Malone, W., & Sweeny, K. (2008). Exploring causes of the self-serving bias. *Social and Personality Psychology Compass*, 2(2), 895–908. <https://doi.org/10.1111/j.1751-9004.2008.00078.x>

Sinclair, L., & Kunda, Z. (2000). Motivated stereotyping of women: She's fine if she praised me but incompetent if she criticized me. *Personality and Social Psychology Bulletin*, 26(11), 1329–1342. <https://doi.org/10.1177/0146167200263002>

Sparqs. (2023, October) Student learning experience model. <https://www.sparqs.ac.uk/resource-item.php?item=293>

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: the state of the art. *Review of Educational Research*, 83(4), 598–642. <https://doi.org/10.3102/0034654313496870>

Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation: A theoretical and empirical analysis. *Basic and Applied Social Psychology*, 42(4), 276–294. <https://doi.org/10.1080/01973533.2020.1756817>

Sultan, P., & Wong, H. Y. (2018). How service quality affects university brand performance, university brand image and behavioural intention: The mediating effects of satisfaction and trust and moderating roles of gender and study mode. *The Journal of Brand Management*, 26(3), 332–347. <https://doi.org/10.1057/s41262-018-0131-3>

Symon, G., Cassell, C., & King, N. (2012). Doing template analysis. In G. Symon & C. Cassell (Eds.), *Qualitative organizational research: Core methods and current challenges* (pp. 426–450). Sage. <https://doi.org/10.4135/9781526435620.n24>

Watty, K. (2006). Want to know about quality in higher education? Ask an Academic. *Quality in Higher Education*, 12(3), 291–301. <https://doi.org/10.1080/13538320601051101>

Willig, C. (2013). *Introducing qualitative research in psychology*. McGraw-Hill.

Wolff, P., & Holmes, K. J. (2011). Linguistic relativity. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(3), 253–265. <https://doi.org/10.1002/wcs.104>

Woodall, T., Hiller, A., & Resnick, S. (2014). Making sense of higher education: Students as consumers and the value of the university experience. *Studies in Higher Education*, 39(1), 48–67. <https://doi.org/10.1080/03075079.2011.648373>

Zabaleta, F. (2007). The use and misuse of student evaluations of teaching. *Teaching in Higher Education*, 12(1), 55–76. <https://doi.org/10.1080/13562510601102131>