



Assessing Student Engagement in Collaborative Learning: Development and Validation of New Measure in China

Bing Xu¹ · Jason M. Stephens¹ · Kerry Lee¹

Accepted: 27 April 2023 / Published online: 19 May 2023
© The Author(s) 2023

Abstract Collaborative learning (CL) is widely used in higher education around the world because it is associated with increases in students' knowledge and social skills. Low student engagement in CL activities has been identified as a common issue, while there is no CL-specific engagement scale to measure and understand engagement in such settings. Additionally, although western countries have been investigating student engagement since the 1950s, there have been comparatively few studies of this important construct in China. To address these imbalances, the present paper adopted a mixed methods approach to developing and validating a Chinese CL engagement scale. In the quantitative study, Chinese university students ($N=405$) completed an anonymous online survey to assess their engagement in CL activities. Results from exploratory and confirmatory factor analyses supported a hierarchical (second-order) three-factor model of student engagement (behavioural, cognitive and emotional engagement), consistent with the tripartite conception of student engagement in Western countries.

Participants were further divided into three unique groups based on engagement scores, and in the qualitative study, 12 participants from three groups were interviewed about CL experiences. Interviews served to further validate the quantitative results. The significance, limitations, and implications of these findings are discussed.

Keywords Student engagement · Collaborative learning · Scale development · Factor analysis

Introduction

Collaborative Learning: The Strengths and Challenges

Collaborative learning (CL) refers to a small group of students performing collective learning activities (e.g., writing a report) with the aim of achieving specific goals (Cohen, 1994; Dillenbourg, 1999). Since CL has been linked with increases in academic achievement (Kollöffel et al., 2011; Le et al., 2018), collaborative and social skills (Ku et al., 2013), and higher-order thinking (Chen et al., 2018), it is used in universities around the world (Echeverria et al., 2022; Johnson & Johnson, 2009). In China, CL is popular not only because of the aforementioned benefits, but also due to the collectivist culture of the country—particularly the belief that people working together can achieve more (or better) than an individual toiling alone (Gong & Cheng, 2020; Phuong-Mai et al., 2005).

Students' CL achievement is associated with their level of engaging (Dillenbourg, 1999; Webb, 2013; Xu et al., 2020), however, disengagement is a common problem with (Le et al., 2018; Sinha et al., 2015; Webb, 2013). Such an issue has become a focus of academic concern and social interest (Jiang, 2011; Xu, 2018). Chen and Bennett (2012) found

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40299-023-00737-x>.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

✉ Bing Xu
bing.xu@auckland.ac.nz

Jason M. Stephens
jm.stephens@auckland.ac.nz

Kerry Lee
k.lee@auckland.ac.nz

¹ Faculty of Education and Social Work, The University of Auckland, 74 Epsom Ave, Auckland 1142, New Zealand

that students complained about the low peer engagement in online discussions and cohesive learning communities were infrequently formed. The official Chinese Youth WeChat Channel (2022) published an article—“What freaks university students out? A group homework will do”—suggesting that group assignments always irritate university students. Followers (mainly university students) commented that their peers’ low engagement was the most difficult problem with CL. Due to the vast number of Chinese university students (around 21 million undergraduate and master students in 2020; National Bureau of Statistics of China, n.d.), this article argues that more attention needs to be placed on their engagement in the widely-used CL.

Student Engagement

Student engagement has been discussed for decades because of its strong association with learning outcomes (Astin, 1984; Kahu, 2013; Skinner & Pitzer, 2012; Trowler, 2010). The present research is rooted in the psychological approach to exploring engagement, particularly Fredricks et al. (2004) tripartite definition that includes: (1) behavioural engagement (BE), which incorporates positive conduct (e.g., following the rules) and involvement in learning and tasks (e.g., effort and attention); (2) cognitive engagement (CE), which focuses on students’ inner psychological investment in learning and being strategic; and (3) emotional engagement (EE), which refers to affective reactions to learning, including students’ academic feelings (e.g., enjoyment, Pekrun et al., 2011), interest, and sense of belonging in the learning community.

According to Skinner and Pitzer (2012), student engagement has been investigated at four nested levels/contexts: *prosocial, school/university, course and learning activity levels*. Although a few studies have examined engagement across different levels and explored the collaborative impact of parents, teachers, and peers (e.g., Skinner et al., 2022; Vollet et al., 2017), a majority of research has focused specifically on the first three levels (e.g., Coates, 2010; Kuh, 2003, 2009). An example of university projects is the China College Student Survey (CCSS). Until 2018, it had assessed over half a million Chinese college students (Huang et al., 2021), and based on such data, researchers identified three factors for student engagement: practical relevance, academic validity, and effective collaborations (Guo et al., 2022).

By comparison, relatively little research on student engagement has been conducted at the learning activity level and even fewer studies discussed how to measure engagement in CL. Prior engagement surveys, for example, School Engagement Survey (Fredricks et al., 2005), Student Engagement Scale (Gunuc & Kuzu, 2015) and Student Engagement Instrument (Appleton et al., 2006) focus

on school/university and course levels. Simply applying the aforementioned macro-level instruments into micro CL settings would only yield a decontextualized understanding of engagement; violating the position that engagement is negotiated within particular contexts and influenced by environmental factors, such as pedagogical practices (Sinha et al., 2015). However, some published surveys inform potential indicators of BE, CE and EE in activity settings, such as paying attention (Gunuc & Kuzu, 2015), adopting meta-cognitive strategies (Muukkonen et al., 2020) and achievement emotions (Pekrun et al., 2005). Additionally, previous surveys have provided a collection of items to the present study.

In summary, two gaps in CL and student engagement have been identified: low engagement in Chinese university CL activities, and a lack of appropriate tools to assess and monitor student CL engagement. To address these gaps, the present investigation adopted a mixed method approach and designed two sequential sub-studies. The quantitative study constructed and validated a new measure of CL engagement in Mandarin. A following qualitative study interviewed a subset of participants from the quantitative study to triangulate the results of the quantitative one. In doing so, this investigation sought to contribute to the CL and engagement fields theoretically and practically.

Quantitative Study

Methods

Participants

In this study, 672 students from 11 courses at six universities were invited to participate. These six universities were located in various regions of China, including the north, west, south and east; three were comprehensive universities and three were polytechnic universities. Out of the invited students, 504 (75.0%) agreed to undertake the survey. However, 99 (19.6%) of those who consented were later removed from the study as they did not complete the survey. The final sample of 405 participants included 331 (81.7%) undergraduate students (249 females and 82 males) and 74 (18.3%) master students (56 females and 18 males). Students were from the faculties of Arts ($n=66$), Business ($n=8$), Education ($n=205$), Engineering ($n=27$) and Science ($n=99$). Finally, two students indicated they were under 18 years old and all others reported being between 18 and 24 years old.

Procedures

Two steps were carried out to identify and recruit student participants in mainland China. The first step involved recruiting university lecturers who were employing CL

activities in line with the criteria of the present study—formal CL assignments that lasted for several weeks. To achieve this goal, the first author published an advertisement introducing the present study's aims on social websites (e.g., Weibo and WeChat). A total of 14 lecturers from 10 universities contacted the first author. After communications about the design and implementation of CL activities, the first author identified that only 11 activities in 11 courses met the criteria.

Secondly, with the consent of lecturers, the first author sent the research advertisement and online anonymous survey links to students in 11 courses less than one month after CL completion. Students had one week to complete the survey. Lecturers agreed that students participating or not would not affect students' grades, as shown in the participant information sheet included in the survey link. Approval for this study was granted by the University of Auckland Human Participants Ethics Committee (Reference Number: UAHPEC3300).

Instrument Development

The instrument was constructed using items from well-established measures of student engagement in other contexts. In keeping with Fredericks et al. (2004), items ($N=46$) were initially organized by the latent construct they were intended to assess, namely BE, CE and EE. See supplementary files (Appendix A) for a complete list of all items and their original sources.

Behavioural engagement. BE was assessed with 8 items from Fredrick et al. (2005) and, Gunuc and Kuzu (2015). Participants were asked to use a five-point Likert-type scale (where 1 = never, 2 = rarely, 3 = sometimes, 4 = often, and 5 = always) to indicate the extent to which they engaged in a range of behaviours during CL (e.g., *I actively interact with peers in my group*).

Cognitive engagement. CE also applied a same five-point Likert-type scale as BE and was evaluated with 18 items from Appleton et al. (2006), Fredrick et al. (2005), Gunuc and Kuzu (2015), Maroco et al. (2016), and Muukkonen et al. (2020). Participants were required to recall how often they exerted psychological energy and applied meta-cognitive strategies to complete CL tasks (e.g., *I try to do my part of groupwork in the best way*).

Emotional engagement. EE contained 16 items from Appleton et al. (2006), Blasco-Arcas et al. (2013), Fredrick et al. (2005), Gunuc and Kuzu (2015), and Pekrun et al. (2005). Items were presented in a nine-point semantic differential scale. Semantic differential scales arrange a set of bi-polar adjectives in pairs (e.g., sad-happy) and participants need to select a point that can best reflect their feelings along the continuum with the nine points, which are assigned scores of -4 to 4 in turn for data analysis.

Semantic differential scales have been used to understand people's emotional responses toward objects, events or concepts (Badia et al., 2014). They can not only gain similar and reliable results as Likert-type scales but also effectively avoid text repetition and redundancy than the latter (Schibeci, 1982). In EE section, participants were asked to report emotional feelings, such as enjoyment and boredom. (e.g., *When I think about our groupwork, I feel bored vs. interested*).

Modifications were made to the original items to fit CL settings. For example, the appellation 'classmate' and context 'classroom/school' were changed to 'peers in group' and 'my group', respectively. The first author together with another two Ph.D. students who were fluent in both English and Mandarin completed translation (from English to Chinese) and back-translation to ensure that the Chinese items were authentic and clear, and expressed the same meaning as the original English ones. The online survey was administered and distributed via Qualtrics.

Data Analysis

Missing values and data analysis assumptions were firstly inspected. In the dataset with 405 cases, 380 (93.8%) participants responded to all survey items ($n=49$, including three demographics items) and 25 (6.2%) participants missed one to five items (questions were not required to be answered for ethical considerations). Specifically, one participant missed five items, one missed four, two missed two items and the remaining 21 participants missed only one response. Among the 19,845 (i.e., 405×49) potential data cells, 34 entries (0.17%) were missed out and the percentage of missing values for each item varied from 0.0 to 1.0%. The expectation-maximization (EM) algorithm was applied to impute missing values (Dong & Peng, 2013; Watkins, 2018). The kurtosis and skewness of each variable were within the normal range for the sample larger than 300 (i.e., $|kurtosis| < 7$, $|skewness| < 2$, Byrne, 2010), indicating the dataset was univariate normal. However, Mardia's (1970) multivariate estimates ($p < 0.05$) showed that the dataset was multivariate non-normal, which further informed the selection of the technique of factor analysis (discussed below). The data were randomly split into two independent samples—one including 200 cases used for exploratory factor analysis (EFA) and another one with 205 cases used for confirmatory factor analysis (CFA). EFA was employed to identify the underlying relationships between the items and CFA to confirm the factor structure suggested by the former result.

EFA was conducted following six steps: (1) Bartlett's sphericity and Kaiser-Meyer-Olkin (KMO) tests; (2) Extracting latent common factors. The common factor analysis model and the iterated principal axis, which has no multivariate normality assumptions, were applied (Cudeck,

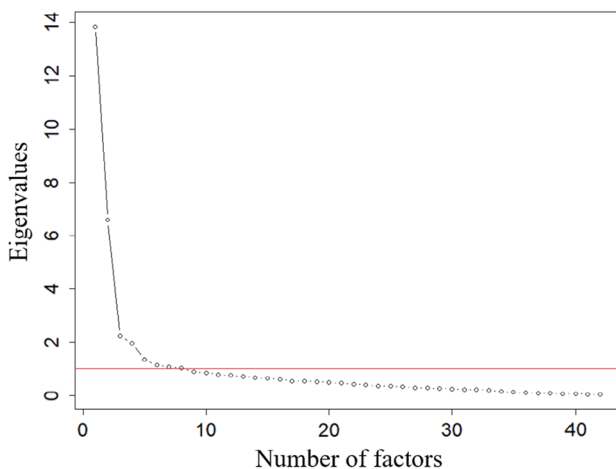
2000); (3) Determining the number of factors by a scree-plot and parallel analysis; (4) Oblimin rotation was used since the latent factors (BE, CE and EE) are known to be correlated (Fredricks et al., 2004); (5) Refining the item-factor structure; and (6) Naming and interpreting latent factors.

In CFA, based on the literature (Hu & Bentler, 1999; Kline, 2016), the following standards were used for a good (or acceptable) fit: the ratio of the chi-square value and the degree of freedom (i.e., $\chi^2/df < 2$ (or 3)); SRMR < 0.05 (or 0.08); RMSEA < 0.05 (or 0.08); CFI and TLI > 0.95 (or 0.90). Data cleaning, assumption checking and EFA were conducted in R® (*MVN* and *psych* packages) and CFA was conducted in AMOS® v. 27.

Results

Exploratory Factor Analysis

The Bartlett's test ($\chi^2 = 6927.70$, $p < 0.05$) and KMO (0.92) showed that variables were sufficiently intercorrelated to each other and supported the appropriateness of EFA. The scree plot result (Fig. 1) indicated five factors since there was no significant drop after the fifth point, while parallel analysis results suggested four factors. Five items (BE_5, BE_7, CE_4, CE_8, and CE_10) were removed since their cross-loading issues may affect the item-factor structure. However, the discrepancy between the two methods still existed after deletion. The factor loading matrices of four and five factors were calculated to decide the number of factors. As detailed in Table 1, compared to the four-factor solution, the five-factor solution contained one fewer item that cross-loaded and better fitness indices (e.g., lower RMSR and higher TLI). Thus, the five-factor model was selected for further testing.



Note. The red line indicates that the eigenvalue equals one.

Fig. 1 The scree plot of eigenvalues

The item-factor loadings in the five-factor model (Table 1) were referenced to decide whether an item should remain. For a sample size of 200, it is recommended that items with a factor loading less than 0.364 or with cross-loading issues (loading differences < 0.05) should be removed (Stevens, 2002). Thus, EE_3 and CE_1 were deleted. In addition, the alpha values also indicated that EE_11 should be deleted for increasing the reliability of Factor 3. On contrary, CE_13 and CE_15 were kept because otherwise, the alpha value of Factor 1 would decrease. Therefore, 34 items remained in the survey after EFA.

Factor 1 and 5 aligned with the hypothesized item-factor structures of CE and BE, respectively. Factors 2 to 4 contained different clusters of EE items, showing there were three sub-factors of EE. According to items' meanings, the definition of EE (Fredricks et al., 2004), and relevant theories (Academic Emotional Theory, Pekrun et al., 2011); Expectancy-value theory, Wigfield and Eccles 2000), Factor 2, Factor 3 and Factor 4 were named as Sense of Belonging (SB), Positive Affective Effect (PA) and Task Value (TV), respectively. In summary, EFA results showed the second-order factor structure of student engagement: SB, TV and PA as the first-order factors of EE, and BE, CE, and EE as the second-order factors of overall student engagement.

Confirmatory Factor Analysis and Descriptive Analysis

The measurement model one (M1) with 34 items in five factors was constructed in AMOS. M1 followed three rules: each item had a nonzero loading on its first-order latent factor and zero loadings on other factors; the factor loadings of the first item in every congeneric were fixed as 1; and error terms of each item were uncorrelated. As presented in Table 2, the indices suggested its poor model fit to the data. The modification indices (MIs) showed three items (i.e., CE_5, CE_13, and EE_2) had high error covariance with one or more other items. Thus, CE_5 was removed in M2, however, the upper limit of 90% confidence interval of RMSEA was higher than the acceptable value. CE_13 was further deleted as the MIs of M2 suggested.

The updated model M3 showed significant improvement compared to M2 according to a Chi-square test ($\Delta\chi^2 = 92.4$, $p < 0.01$). Although all indices of M3 were in their acceptable ranges, MIs still suggested to remove EE_2 because it was highly correlated with EE_1. M4 with EE_2 removed was significantly better than M3 ($\Delta\chi^2 = 197.8$, $p < 0.01$) and MIs did not indicate any further modifications. Therefore, M4 was selected as the most appropriate measurement model, and its standardized path estimates are presented in Fig. 2. A multiple-group CFA showed M4 to be equivalent at the configural (CFI = 0.93, TLI = 0.93, RMSEA = 0.05), metric ($\Delta\chi^2_{(26)} = 29.14$, $p > 0.05$, $\Delta CFI = 0.001$) and scalar levels ($\Delta\chi^2_{(28)} = 33.62$, $p > 0.05$, $\Delta CFI = 0.001$), indicating

Table 1 Five-factor and four-factor loading matrices estimated by the iterated principal axis (PA) method

Item #	Five-factor loading matrix					Item #	Four-factor loading matrix			
	F1	F2	F3	F4	F5		F1	F2	F3	F4
CE_16	.72					CE_7	.64			
CE_18	.67					CE_6	.41			.43
CE_5	.64					CE_12	.58			
CE_7	.63					CE_2	.34			.41
CE_9	.63					BE_2	.34			.44
CE_12	.58					BE_6	.31			.42
CE_3	.56	.32				CE_1	.36			.38
CE_11	.53					CE_17	.50			
CE_17	.49					CE_16	.73			
CE_14	.50					CE_9	.63			
CE_15	.48					CE_14	.49			
CE_13	.44					CE_13	.46			
EE_15		.96				CE_5	.66			
EE_12		.90				CE_11	.54			
EE_13		.88				BE_8				.78
EE_16		.84				BE_1				.55
EE_14		.86				CE_15	.50			
EE_7			.70			BE_4				.64
EE_6			.69			CE_3	.57	.34		
EE_2			.88			CE_18	.69			
EE_1			.90			BE_3				.60
EE_5			.65			EE_15	.92			
EE_4			.49			EE_12	.94			
EE_3			.34			EE_13	.92			
EE_10				.97		EE_16	.89			
EE_9				.92		EE_14	.83			
EE_8				.81		EE_7			.74	
EE_11				.61		EE_6			.77	
BE_8					.79	EE_2			.86	
BE_4					.65	EE_1			.91	
BE_3					.59	EE_5			.86	
BE_1					.55	EE_4			.54	
CE_6	.39				.44	EE_3			.33	
CE_2	.30				.44	EE_10	.72			
BE_6	.34				.43	EE_9	.74			
BE_2	.31				.43	EE_8	.70			
CE_1	.35				.39	EE_11	.50		.39	
Sum of squared loading	5.47	5.15	4.37	3.83	3.52		7.28	5.64	4.93	3.44
RMSR	.03					.04				
TLI	.907					.796				

The loadings less than .3 were omitted in this table. F= factor

the overrepresentation of females in the sample did not affect the instrument validity.

Descriptive analysis and correlation analysis of five latent factors were conducted (see Table 3). BE and CE were mediumly correlated to each other ($r=0.61^{**}$), while both were

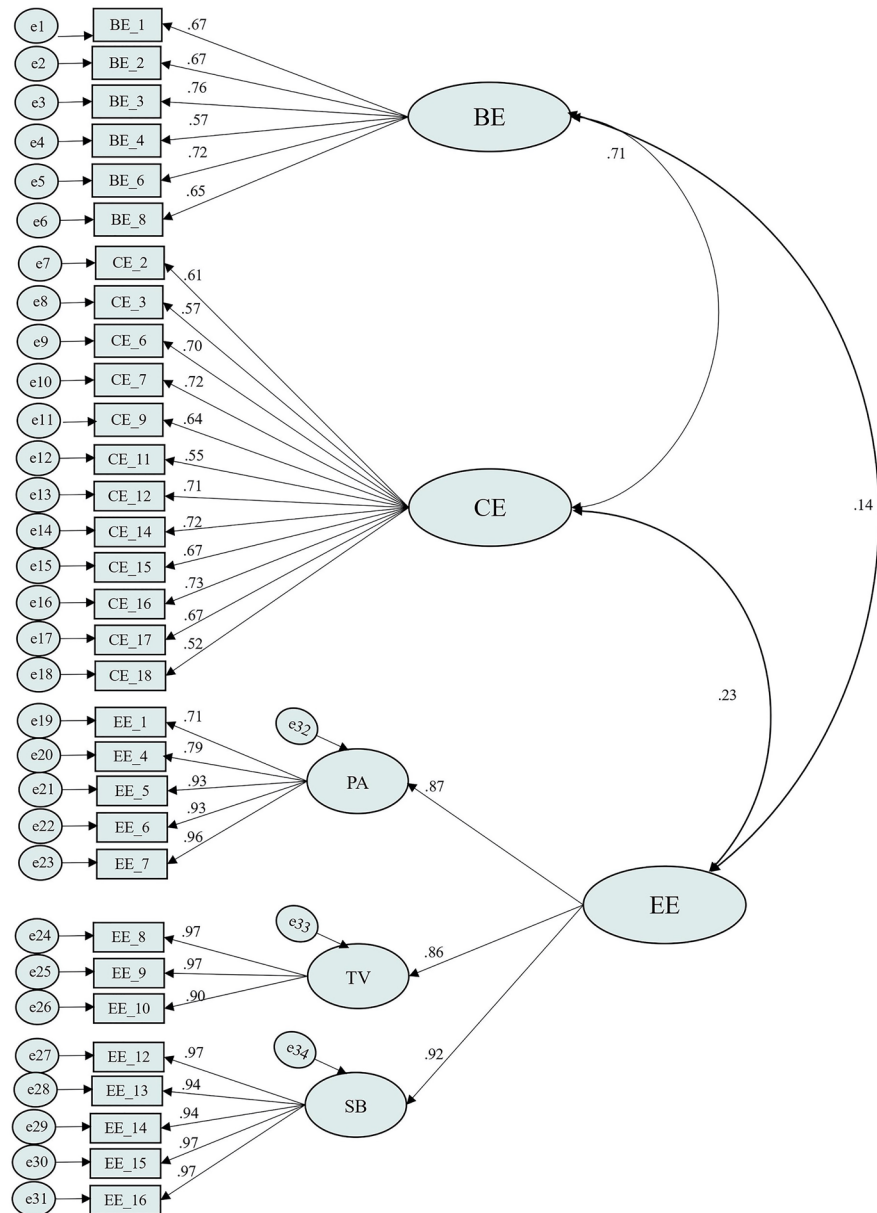
slightly associated with EE and three sub-factors of EE ($0.16^{**} \leq r \leq 0.29^{**}$). PA, TV and SB were mediumly correlated to each other, and the high path estimate (see Fig. 2) between SB and EE indicated the importance of SB among

Table 2 Goodness-of-fit indices for measurement models of student engagement

Model	Number of items	χ^2 (df)	χ^2/df	CFI	TLI	RMSEA (90%CI)	SRMR	AIC
M1	34	1159.68 (521)	2.23	.898	.890	.078 (.072, .084)	.059	1307.68
M2	33	1052.37 (489)	2.15	.908	.901	.075 (.069, .081)	.057	1196.37
M3	32	959.97 (458)	2.10	.916	.909	.073 (.067, .080)	.057	1099.97
M4	31	762.17 (428)	1.78	.941	.936	.062 (.055, .069)	.054	898.12

N=405. M1 = Model 1; M2 = M1 – item CE_5; M3 = M2—item CE_13; and M4 = M3—item EE_2

Fig. 2 Multifactor student engagement model



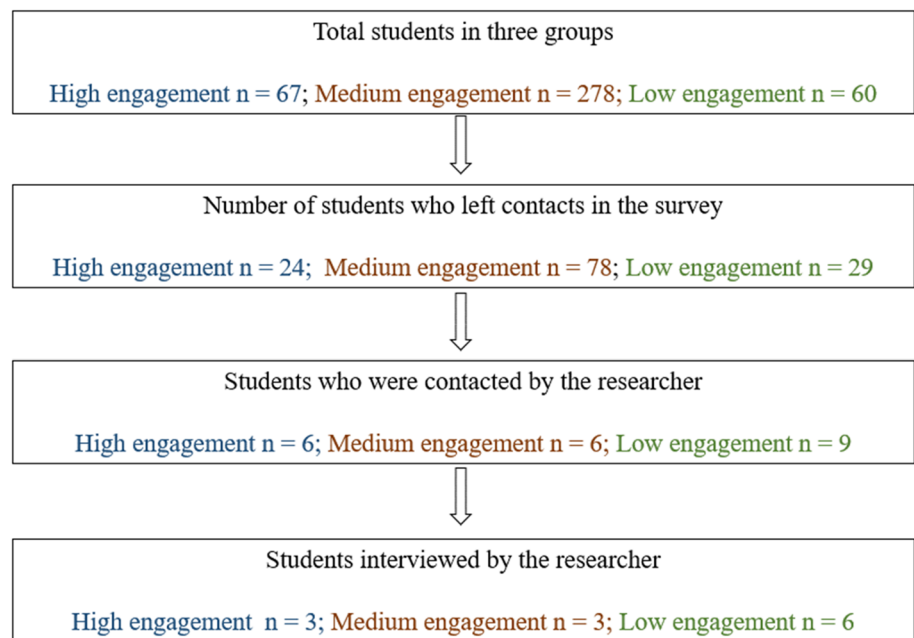
Note. BE = Behavioural engagement, CE = Cognitive engagement, EE = Emotional engagement, PA = Positive affect, TV = Task value, SB = Sense of belonging. Path estimates were standardized.

Table 3 Descriptive analysis and correlation of latent factors of student engagement

Factors	Number of Items	M	SD	Actual range	α	Correlation				
						BE	CE	EE	PA	TV
BE	6	4.28	0.52	1.67~5	.82					
CE	12	3.78	0.59	2.25~5	.88	.62**				
EE	13	1.14	1.84	-4~4	.96	.21**	.28**			
- PA	5	0.73	1.99	-4~4	.94	.15**	.26**	.90**		
- TV	3	1.50	2.17	-4~4	.97	.16**	.21**	.88**	.71**	
- SB	5	1.33	2.07	-4~4	.98	.25**	.29**	.92**	.72**	.76**

N=405. BE=Behavioural engagement, CE=Cognitive engagement, EE=Emotional engagement, PA=Positive affect, TV=Task value, SB=Sense of belonging. **p*<.05, ***p*<.01

Fig. 3 The process of recruiting participants for the qualitative study



the three EE sub-factors. The alpha values indicated high reliabilities of sub-scales and the whole scale ($\alpha=0.93$).

Qualitative Study

Methods

Procedures and Participants

Figure 3 depicts the participants recruiting process of this qualitative study. Firstly, participants from the quantitative study were divided into three groups based on their overall scores: 1) high engagement (overall score > $M + 1SD$; $n = 67$), 2) medium engagement ($M - 1SD \leq$ overall score $\leq M + 1SD$; $n = 278$), and 3) low engagement (overall score < $M - 1SD$; $n = 60$). Secondly, the first author

contacted 21 students who indicated (at the end of the survey) interest in a follow-up interview while 9 rejected the invitation, therefore, 12 students ($N = 12$) were interviewed. The sample size was determined by data saturation—the point, at which, further data collection was no longer adding value or new insights (Braun & Clark, 2013). Interviewees varied in engagement levels and demographics (Table 4). The author intentionally over-sampled the low engagement group as the feedback from these students’ will be utilized in another paper investigating influencing factors of CL disengagement.

Interviews

To understand participants’ experiences and engagement during CL tasks, the interviewer conducted semi-structured interviews after building rapport with interviewees (see Appendix B for guideline questions). All interviews were

Table 4 Demographics of participants in the qualitative study

Pseudonyms	Gender	Level-Year	Faculty
High engagement			
Diana	F	Undergraduate-Y2	Education
Faye	F	Undergraduate-Y3	Science
Penny	F	Master-Y1	Education
Medium engagement			
Mia	F	Undergraduate-Y3	Education
Xavier	M	Undergraduate-Y3	Engineering
Perl	F	Master-Y1	Education
Low engagement			
Helen	F	Undergraduate-Y1	Arts
Sean	M	Undergraduate-Y2	Science
Zac	M	Undergraduate-Y2	Science
William	M	Undergraduate-Y2	Engineering
Leo	M	Undergraduate-Y3	Science
Joan	F	Master-Y1	Education

F = female, M = male; Y = Year

conducted online via WeChat and audio-recorded. Each interview lasted for 40 min on average. The audio recordings were transcribed into text via the website [iflytek](#) and further edited by the first author. The transcripts were sent to participants to check the accuracy, trustworthiness, and authenticity.

Results

In general, students' opinions and perceptions towards their engagement in CL tasks aligned with the quantitative study. Students from the high engagement group, Diana, Faye and Penny recognized they were behaviourally and cognitively engaged in CL. Faye stated that "I did the literature review and worked with others on our presentation slides... I tried my best to do what I can do". Diana also mentioned that she made a large contribution in the group work and felt satisfied with her engagement as "I would score myself 90 out of 100". Three participants also expressed their favour towards the task and their collaborators (i.e., emotional engagement). As Penny (in the course of Instruction design and implementation) said:

I like my task. I love exploring students' needs and designing something [e.g., courses] new for them... We need to restructure teaching content because the well-organized teaching content benefits students.

The medium engagement group (Mia, Xavier and Perl) was pleased with their contributions, yet acknowledged the gap between what they could have achieved and what they did achieve. This self-evaluation verified their survey results. Mia stated that "I have completed part of the group work, like searching resources and organizing key ideas ...

However, in my opinion, our group product and what I have done, were just okay, not very outstanding". Likewise, Perl acknowledged that the outcomes did not reach her expectation. Xavier (in the Wine experiment course) admitted he did not engage in the CL task deeply and explained the reasons:

I know these experiments are essential for wine production. But these skills are necessary for the people who plan to be inspectors. Well, I don't want to do that in the future. So I just ensured I can understand the experiments, rather than setting a high level for myself in this course.

The other six students recognized that they did not engage in the CL task which triangulated the low survey scores they received. For example, Leo admitted: "I can't say I contributed knowledge or good ideas to my group. In fact, I almost disengaged". Some students also explained that the group members did not have adequate collaborative competencies to interact deeply and work smoothly. As Helen stated, "My peers and I were so fresh as first-year students. Collaborative learning was very strange to us, so we didn't collaborate very well".

Discussion

CL is widely spreading around the world. Students' achievements in CL activities are dependent on effective engagement. Low engagement has been identified as a common issue while it has been paid relatively little attention in China. This paper situated student engagement in the Chinese CL context and developed a Chinese engagement scale. The results are discussed below.

The Collaborative Learning Engagement Scale and Engagement Construct

The quantitative study constructed and validated a Mandarin CL engagement scale, and explored the conceptual structure of engagement. Results from EFA and CFA showed the high reliability of the scale and indicated a hierarchical structure of engagement, with BE, CE and EE being the second-order factors, and PA, TV and SB being the first-order factors of EE. BE measured participants' positive conduct and involvement in CL, including both individual effort (e.g., BE_4) and group commitment (e.g., BE_1) indicators, which have not been clarified in previous engagement literature. CE was about students' cognitive effort, meta-cognitive strategies (e.g., shared regulation) and the motivation to complete CL tasks. As for measuring EE, the semantic differential scale was creatively used in the present study and it gained reliable results.

The first-order factors of EE, including PA, TV and SB, perfectly aligned with the core elements in Fredrick

et al.'s (2004) definition, however, these three sub-factors were not overtly identified by prior engagement scales. For example, scales may incorporate Teacher-Student Relationships (Appleton et al., 2008), Peer Relationships, and Relationships with the Faculty Member (Gunuc & Kuzu, 2015), in the emotional dimension. The present investigation suggested that students' perceptions towards tasks (e.g., PA and TV) and teams (SB) were critical indicators. Such differences were reasonable because previous studies were conducted at the course or school level, while the present study focused on learning activities, where students' feelings towards tasks and collaborators were more directly influencing learning process and outcomes (Pekrun & Linnenbrink-Garcia, 2012).

The qualitative study interviewed 12 students, who reported high, medium and low levels of engagement in the scale, and triangulated the quantitative results. The high-level engagement interviewees were satisfied with their contribution and engagement; the medium group recognized their contribution, yet acknowledged that they could have engaged better; the low engagement group admitted that they hardly contributed or engaged in CL.

Furthermore, this paper found BE and CE were moderately correlated with each other, mirroring previous findings at both activity (Naibert and Barbera 2022) and school (Virtanen et al. 2018) levels. Such results indicated that participants would more likely engage with behaviours and cognition synchronously, and students' effort, attention and concentration (i.e., BE) were important in facilitating CE (Wilson et al. 2021; Wolters and Hussain 2015). The present investigation also found that EE had low correlation with BE and CE. It was not consistent with Pekrun's control-value theory which proposed when students have positive sense towards a task, they will be willing to exert more time and effort on it (Pekrun et al. 2002; Pekrun & Linnenbrink-Garcia, 2012). However, the low correlation between EE and BE/CE was also found in a longitudinal study conducted by Manwaring et al. (2017) in higher education. Such results indicated that some adult learners were less dependent on their emotional feelings (e.g., interest in tasks) to engage behaviourally and cognitively, but engaging for external rewards. Nonetheless, other students still attributed their investment (or disengagement) to interest in (or dislike for) the task and the collaborating experiences with peers.

Limitations and Future Research

Several limitations can be found in the present research. This study was cross-sectional observation research, therefore, the results cannot make causal conclusions. Additionally, most participants were female from humanities majors, which may not be able to represent all university students in China. In addition, the survey was designed to

measure overall engagement levels post hoc. However, student engagement is dynamic and can vary from time to time during CL. Accordingly, future researchers are encouraged to conduct longitudinal research and adopt micro-genesis methods to capture finer-gained learning data. Further studies could also explore what factors influence student engagement in CL and how.

Implications and Significance of the Present Research

This research contributes to the realm of student engagement and CL practically and theoretically. The validated scale can serve as an instrument in future research that focuses on CL or student engagement in other learning ac and applied in other language contexts. Teachers could also use the full scale (or one of the subscales) to assess and monitor student engagement, and intervene accordingly as needed. Moreover, the refined conceptual structure and clarified indicators can guide CL design and assessment. For example, since BE covers indicators at both individual and group levels, teachers are suggested to assess CL tasks at two levels as well. Considering PA, SB and TV, teachers could (1) ensure tasks align with students' intrinsic interests, (2) help build up team cohesion and interdependence, and (3) clarify the values and benefits of fully engaging in CL, to enhance students' EE. Only with effective actions being taken appropriately, can the low engagement issues be solved and the expected outcomes of CL be achieved.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Declarations

Conflict of Interest No competing interests have been reported by the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Appleton, J. J., Christenson, S. L., Kim, D., & Reschly, A. L. (2006). Measuring cognitive and psychological engagement: Validation of

- the student engagement instrument. *Journal of School Psychology*, 44(5), 427–445. <https://doi.org/10.1016/j.jsp.2006.04.002>
- Appleton, J. J., Christenson, S. L., & Furlong, M. J. (2008). Student engagement with school: critical conceptual and methodological issues of the construct. *Psychology in the Schools*, 45(5), 369–386. <https://doi.org/10.1002/pits.20303>
- Astin, A. (1984). Student involvement: a development theory for higher education. *Journal of College Student Development*, 25(4), 297–308.
- Badia, A., Meneses, J., & Monereo, C. (2014). Affective dimension of university professors about their teaching: An exploration through the semantic differential technique. *Universitas Psychologica*, 13(1), 161–174.
- Blasco-Arcas, L., Buil, I., Hernández-Ortega, B., & Sese, F. J. (2013). Using clickers in class. The role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, 62, 102–110. <https://doi.org/10.1016/j.compedu.2012.10.019>
- Chen, R. T.-H., & Bennett, S. (2012). When Chinese learners meet constructivist pedagogy online. *Higher Education*, 64(5), 677–691.
- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C.-C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: a meta-analysis. *Review of Educational Research*, 88(6), 799–843. <https://doi.org/10.3102/0034654318791584>
- Coates, H. (2010). Development of the Australasian survey of student engagement (AUSSE). *Higher Education*, 60(1), 1–17.
- Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research*, 64(1), 1–35. <https://doi.org/10.2307/1170744>
- Cudeck, R. (2000). Exploratory factor analysis. In H. E. A. Tinsley & S. D. Brown (Eds.), *Handbook of applied multivariate statistics and mathematical modeling* (pp. 265–296). Elsevier Academic Press.
- Dillenbourg, P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: cognitive and computational approaches* (pp. 1–19). Elsevier.
- Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *Springerplus*, 2(1), 222. <https://doi.org/10.1186/2193-1801-2-222>
- Echeverria, V., Martinez-Maldonado, R., Yan, L., Zhao, L., Fernandez-Nieto, G., Gasevic, D., & Shum, S. B. (2022). HuCETA: a framework for human-centered embodied teamwork analytics. *IEEE Pervasive Computing*. <https://doi.org/10.1109/MPRV.2022.3217454>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59–109. <https://doi.org/10.3102/00346543074001059>
- Fredricks, J. A., Blumenfeld, P., Friedel, J., & Paris, A. (2005). School engagement. In K. A. Moore & L. H. Lippman (Eds.), *What do children need to flourish? Conceptualizing and measuring indicators of positive development* (pp. 305–321). Springer.
- Gong, B., & Cheng, T. (2020). Hezuoxuexi: Qingmian, hehe yu jiaoxuezhixun zaishengchan-jiyu guocheng jiegou shijiao de bijiao gean fenxi (Cooperative learning: Qingmian, hehe and teaching order reproduction). *Gaodeng jiaoyu yanjiu*, 41(2), 66–75. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFQ&dbname=CJFDLAST2020&filename=HIGH202002009&v=MzAwMTBITXJZOUZiWVI4ZVgxTHV4WVM3RGxVDNkxVHJXTTFGckNVUjdxZlIPVnZGaS9oVUx2SkxTVE1ackc0SE4=>
- Gunuc, S., & Kuzu, A. (2015). Student engagement scale: development, reliability and validity. *Assessment & Evaluation in Higher Education*, 40(4), 587–610. <https://doi.org/10.1080/02602938.2014.938019>
- Guo, F., Gao, X., Yang, J., & Shi, J. (2022). CCSS and student engagement in China: From an academic concept to institutional practices and policy implications. In H. Coates, X. Gao, F. Guo, & J. Shi (Eds.), *Global student engagement* (pp. 27–42). Routledge.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Huang, Y., Zhou, X., & Shi, J. (2021). Woguo benke kecheng jiaoxue zhiliang zhenmeyang? Jiyu zhongguo daxuesheng xuexi yu fazhan zhuzhong yanjiu de shinian tansuo [The quality of undergraduate teaching and learning in China: a ten-year exploration based on China College Student Survey]. *Huadong Shifan Daxue Xuebao (jiaoyu Kexue Ban)*, 1, 116–126.
- Jiang, B. (2011). Hezuoxuexi dixiaoxing de duwei yinsu fenxi [Exploring the reasons of low effectiveness of collaboration learning]. *Jiaoyu Xueshu Yuekan*, 08, 98–110.
- Johnson, D. W., & Johnson, R. T. (2009). An educational psychology success story: Social interdependence theory and cooperative learning. *Educational Researcher*, 38(5), 365–379. <https://doi.org/10.3102/0013189X09339057>
- Kahu, E. R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, 38(5), 758–773. <https://doi.org/10.1080/03075079.2011.598505>
- Kline, R. B. (2016). *Principles and Practice of Structural Equation Modeling* (4th ed.). Guilford Publications.
- Kollöffel, B., Eysink, T. H. S., & Jong, T. de. (2011). Comparing the effects of representational tools in collaborative and individual inquiry learning. *International Journal of Computer-Supported Collaborative Learning*, 6, 223–251. <https://doi.org/10.1007/s11412-011-9110-3>
- Ku, H.-Y., Tseng, H. W., & Akarasriworn, C. (2013). Collaboration factors, teamwork satisfaction, and student attitudes toward online collaborative learning. *Computers in Human Behavior*, 29(3), 922–929. <https://doi.org/10.1016/j.chb.2012.12.019>
- Kuh, G. D. (2003). What we're learning about student engagement from NSSE: benchmarks for effective educational practices. *Change Magazine of Higher Learning*. <https://doi.org/10.1080/00091380309604090>
- Kuh, G. D. (2009). What student affairs professionals need to know about student engagement. *Journal of College Student Development*, 50(6), 683–706. <https://doi.org/10.1353/csd.0.0099>
- Le, H., Janssen, J., & Wubbels, T. (2018). Collaborative learning practices: teacher and student perceived obstacles to effective student collaboration. *Cambridge Journal of Education*, 48(1), 103–122. <https://doi.org/10.1080/0305764X.2016.1259389>
- Manwaring, K. C., Larsen, R., Graham, C. R., Henrie, C. R., & Halverson, L. R. (2017). Investigating student engagement in blended learning settings using experience sampling and structural equation modeling. *The Internet and Higher Education*, 35, 21–33. <https://doi.org/10.1016/j.iheduc.2017.06.002>
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.2307/2334770>
- Maroco, J., Maroco, A. L., Campos, J. A. D. B., & Fredricks, J. A. (2016). *University student's engagement: Development of the University Student Engagement Inventory (USEI)*. *Psicologia: Reflexão e Crítica*. <https://doi.org/10.1186/s41155-016-0042-8>
- Muukkonen, H., Lakkala, M., Lahti-Nuuttila, P., Ilomäki, L., Karlgren, K., & Toom, A. (2020). Assessing the development of collaborative knowledge work competence: scales for higher education course contexts. *Scandinavian Journal of Educational Research*, 64(7), 1071–1089. <https://doi.org/10.1080/00313831.2019.1647284>

- Naibert, N., & Barbera, J. (2022). Investigating student engagement in general chemistry active learning activities using the activity engagement survey (AcES). *Journal of Chemical Education*, 99(7), 2620–2629. <https://doi.org/10.1021/acs.jchemed.2c00276>
- National Bureau of Statistics of China. (n.d.). *National Data*. National Data. <https://data.stats.gov.cn/english/easyquery.htm?cn=C01>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: the achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). Springer. https://doi.org/10.1007/978-1-4614-2018-7_12
- Pekrun, R., Goetz, T., & Perry, R. P. (2005). *Achievement emotions questionnaire (AEQ)—User's manual*. <https://www.researchgate.net/profile/Javier-Sanchez-Rosas/post/How-can-i-apply-and-access-the-Academic-Emotions-Questionnaire-by-Pekrun-et-al-2005/attachment/59d6351c79197b8077992b77/AS%3A383031052390406%401468333127197/download/2005+AEQ+Manual.pdf>
- Phuong-Mai, N., Terlouw, C., & Pilot, A. (2005). Cooperative learning vs Confucian heritage culture's collectivism: confrontation to reveal some cultural conflicts and mismatch. *Asia Europe Journal*, 3(3), 403–419. <https://doi.org/10.1007/s10308-005-0008-4>
- Schibeci, R. A. (1982). Measuring student attitudes: Semantic differential or Likert instruments? *Science Education*, 66(4), 565–570. <https://doi.org/10.1002/sce.3730660407>
- Sinha, S., Rogat, T. K., Adams-Wiggins, K. R., & Hmelo-Silver, C. E. (2015). Collaborative group engagement in a computer-supported inquiry learning environment. *International Journal of Computer-Supported Collaborative Learning*, 10(3), 273–307. <https://doi.org/10.1007/s11412-015-9218-y>
- Skinner, E. A., Rickert, N. P., Vollet, J. W., & Kindermann, T. A. (2022). The complex social ecology of academic development: a bioecological framework and illustration examining the collective effects of parents, teachers, and peers on student engagement. *Educational Psychologist*, 57(2), 87–113. <https://doi.org/10.1080/00461520.2022.2038603>
- Skinner, E. A., & Pitzer, J. R. (2012). Developmental dynamics of student engagement, coping, and everyday resilience. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 21–44). Springer US. https://doi.org/10.1007/978-1-4614-2018-7_2
- Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Routledge. <https://doi.org/10.4324/9781315814919>
- Trowler, V. (2010). *Student engagement literature review*. The Higher Education Academy.
- Virtanen, T. E., Lerkkanen, M.-K., Poikkeus, A.-M., & Kuorelahti, M. (2018). Student engagement and school burnout in Finnish lower-secondary schools: Latent profile analysis. *Scandinavian Journal of Educational Research*, 62(4), 519–537. <https://doi.org/10.1080/00313831.2016.1258669>
- Vollet, J. W., Kindermann, T. A., & Skinner, E. A. (2017). In peer matters, teachers matter: peer group influences on students' engagement depend on teacher involvement. *Journal of Educational Psychology*, 109(5), 635–652. <https://doi.org/10.1037/edu0000172>
- Watkins, M. W. (2018). Exploratory factor analysis: a guide to best practice. *Journal of Black Psychology*, 44(3), 219–246. <https://doi.org/10.1177/0095798418771807>
- Webb, N. (2013). Information processing approaches to collaborative learning. In C. E. Hmelo-Silver, C. A. Chinn, C. K. Chan, & A. M. O'Donnell (Eds.), *The international handbook of collaborative learning* (pp. 19–40). Routledge.
- Wigfield, A., & Eccles, J. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <https://doi.org/10.1006/ceps.1999.1015>
- Wilson, D., Wright, J., & Summers, L. (2021). Mapping patterns of student engagement using cluster analysis. *Journal for STEM Education Research*, 4(2), 217–239. <https://doi.org/10.1007/s41979-021-00049-z>
- Wolters, C. A., & Hussain, M. (2015). Investigating grit and its relations with college students' self-regulated learning and academic achievement. *Metacognition and Learning*, 10(3), 293–311. <https://doi.org/10.1007/s11409-014-9128-9>
- Xu, B., Chen, N.-S., & Chen, G. (2020). Effects of teacher role on student engagement in WeChat-Based online discussion learning. *Computers & Education*, 157, 1–11. <https://doi.org/10.1016/j.compedu.2020.103956>
- Xu, H. (2018). Daxuesheng hezuoxuexi yingxiangyinsu de zhixingyanjiu [The influencing factors of college students collaborative learning: A qualitative approach]. *Gaojiao tansuo*, 08, 44–49. <https://kns.cnki.net/kcms/detail/detail.aspx?dbcode=CJFD&dbname=CJFDLAST2018&filename=GJTA201808008&uniplatform=NZKPT&v=0b-hdufpUSNC7WIBrMdy-asJxZ4e-ff6sTDSXVrSLQSmPa9hyfMjyUFcZrvo45R>
- Youth Beijing. (2022, May). *Qiku daxuesheng de, zhixuyao yifen xiaozu zuoye [What freaks university students out? A group homework will do]*. Chinese Youth WeChat Channel. http://mp.weixin.qq.com/s?__biz=MzA5MjlxOTYxMA==&mid=2653579903&idx=1&sn=0d65b1449ffb469528acad6be5bf77&chksm=8baebcbabcd932dc238cf4e408bfa0192c65abf6a0ac330e3bb41aa22d1efe120bc6def8f78e#rd

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