

Perception of Workload and its Relation to Perceived Teaching and Learning Environments among Finnish and Chinese University Students

Elina Hernesniemi¹, Hannu R äy¹, Kati Kasanen¹, Xuejiao Cheng², Jianzhong Hong³ & Matti Kuittinen¹

¹ University of Eastern Finland, Joensuu, Finland

² Hubei University of Medicine, Shiyang, China

³ Central China Normal University, Wuhan, China

Correspondence: Elina Hernesniemi, School of Educational Sciences and Psychology, University of Eastern Finland, P. O. Box 111, FIN-80101 Joensuu, Finland.

Received: August 18, 2017

Accepted: September 6, 2017

Online Published: September 12, 2017

doi:10.5430/ijhe.v6n5p42

URL: <https://doi.org/10.5430/ijhe.v6n5p42>

Abstract

Earlier research has shown that an excessive workload has a substantial negative effect on students' well being. The present study examines how Finnish and Chinese university students' perceptions of workload are related to their perceptions of the teaching and learning environments (TLE). This is done in order to determine whether the perceived workload of students could be reduced by improving the quality of their TLE. Also the levels of experienced workload of Finnish and Chinese students are compared in this study. The group of participants consisted of 3035 Finnish students and 2309 Chinese students. Since this work is cross-cultural in nature, the adequacy of the level of structural equivalence of the research instrument was confirmed, and, when appropriate, the effect of different response styles on the results was taken into account. Both standard and robust statistical methods were used for the analyses. The results show that in both the Finnish and the Chinese groups the students' perceptions of their workload and TLE are significantly but rather weakly related. Furthermore, irrespectively of whether the response styles are accounted for or not, the results indicate that Chinese students perceive a heavier workload than do Finnish students.

Keywords: Workload, Teaching-learning environment, University students, Cross-cultural study

1. Introduction

Students' perception of their workload has a major effect on their well-being and study success. Indeed, excessive workload has been associated with many problems, such as decreased performance and motivation to study, burnout, anxiety, and depression (Bachman & Bachman, 2006; Jacobs & Dodd, 2003; Parkinson, Gilling, & Suddaby, 2006). A student's workload could be interpreted as the number of hours actually worked, but students' perception of their workload is, in fact, only weakly related to the number of hours used for studying (Kember & Leung, 1998). Therefore it is the student's perception of the workload that is recorded, and a recognition of the means to reduce students' perceived workload is valuable when the quality of education is improved.

In this work the goal is to compare the levels of experienced workload between Finnish and Chinese students and to examine the relationships between perceived teaching and learning environment (TLE) and perceived workload among students in Finland and in mainland China. Also gender differences within the cultures are studied. Finland and China have rather different cultural and educational backgrounds: in Finland the university system originates from the Humboldtian model (Davies, Weko, Kim, & Thulstrup, 2009), whereas the basis of Chinese education lies in the Confucian tradition (Gu, 2006). Despite their differences both systems have fared very well in cross-national comparisons (<http://www.oecd.org/pisa/>), which makes studying the Finnish and Chinese present university student cohorts interesting.

When researching university students in Hong Kong, Kember and Leung noted that besides the content being taught and its difficulty, also students' perceptions of TLE affect their perceptions of workload (Kember, 2004; Kember & Leung, 2006). This led Kember and Leung to conclude that it is possible to inspire students to work long hours, without them perceiving their workload as being excessive, if attention is paid to their TLE. Specifically, Kember (2004) found that good student-student relationships and good teacher-student relationships lead to students' perceiving a lighter workload, whereas frequent examinations lead students' perceiving a heavier workload. Later

results obtained with quantitative examinations (Kember & Leung, 2006) confirmed the existence of a relationship between perceived TLE and perceived workload among students in Hong Kong.

The educational system in Hong Kong, however, differs considerably from the educational system in mainland China, since it has been strongly influenced by the British system (Watkins, 2003). This is especially true at the university level, where many of the teachers are western expatriates (Watkins, 2003). Therefore, it is possible that the results obtained from students in Hong Kong by Kember and Leung may be directly generalizable to neither mainland Chinese nor Western educational systems. The present study attempts to address this problem.

This study focuses on six elements of TLE, i.e., **teaching quality**, **student relations**, **encouraging teachers**, **helpful evaluation**, **teaching for deep understanding**, and **research-based teaching**, and examines two research questions:

- Do the levels of perceived **suitable workload** differ between Finnish and Chinese students or between female and male groups within the cultures?
- How are Finnish and Chinese female and male students' perceptions of their TLE related to their perceptions of **suitable workload**?

As a research assumption, relationships are presumed to exist between all six elements of TLE studied here and students' perception of **suitable workload**. Based on previous results of Kember and Leung (Kember, 2004; Kember & Leung, 2006), the first five TLE elements (in the list above) can be assumed to be associated with perceived workload. The last element, **research-based teaching**, is presumed to be related to **suitable workload**, since connecting teaching and disciplinary research may require students to participate more actively in class (Griffits, 2004), which, in turn, may affect the students' perception of their workload. No specific assumptions about the similarity or difference in the levels of experienced workload between Finnish and Chinese students or genders can be made based on previous results, since this work is to our knowledge the first time when this kind of comparison of these two cultural groups is made.

2. Material and Methods

2.1 Data Collection

The group of participants in this study consisted of 3035 Finnish students at nine Finnish universities and 2309 Chinese students at seven universities in Wuhan city, Central China. Care was taken that in both countries the most popular academic disciplines were included in the sample. In Finland study programs were randomly selected from the largest schools and faculties of all but one of the major Finnish universities. Students' email addresses were then obtained from the Student Registry Offices. In the spring of 2009 17 992 Finnish students received an invitation to complete a web-based questionnaire, which yielded a 17% response rate after three mailing rounds. In China all the 211-program key universities in Wuhan city were represented in the Chinese sample and a random selection from an exhaustive list of both business-oriented and non-business-oriented study programs was made (the 211-project is a program of the Chinese government in an effort to enable the top ca. 100 national universities to become world class universities in the 21st century, see Gallagher et al., 2009). In the summer of 2011, after obtaining permits from the teachers and the students of the selected study programs, 2402 students were asked to complete the paper surveys in their classes during lectures and the questionnaires were then collected on the spot. A 96% response rate was reached with this procedure.

Table 1. Characteristics of Finnish and Chinese students.

		Finnish students		Chinese students	
		n	%	n	%
Year of study	1st	980	32.3	685	29.7
	2nd	942	31.0	383	16.6
	3rd	554	18.3	435	18.8
	4th or more	559	18.4	746	32.3
	Unknown			60	2.6
Gender	Female	2431	80.1	1064	46.1
	Male	604	19.9	1189	51.5
	Unknown			56	2.4
Faculty	Agriculture or Forestry	133	4.4		
	Arts	637	21.0		
	Behavioral Sciences	582	19.2		
	Biological or Environmental Sciences	111	3.7		
	Economics	54	1.8		
	Law	177	5.8		
	Medicine	346	11.4		
	Pharmacy	77	2.5		
	Science	223	7.3		
	Social Sciences	458	15.1		
	Technology	111	3.7		
	Theology	126	4.1		
University ^a	Huazhong Agricultural University			202	8.4
	Zhongnan University of Economics and Law			461	19.1
	Wuhan University of Technology			413	17.1
	Huazhong University of Science and Technology			370	15.3
	China University of Geosciences			419	17.4
	Wuhan University (Chinese literature, History, etc.)			208	8.6
	Central China Normal University (Psychology, Rural and regional development, etc.)			339	14.1

a. Here n and % refer to the number of questionnaires distributed.

The sample consisted of both graduate and undergraduate students, with foreign and exchange students excluded from the study. The mean ages of the Finnish and the Chinese students were 25 ($SD=7.4$) and 21 ($SD=2.3$) years, respectively. The median age was 22 years for the Finnish students and 21 years for the Chinese students. Additional characteristics of the participants are presented in Table 1.

In the Finnish sample 80% were women, which is a disproportionate number, as only 54% of Finnish university students were women in 2009 when the data was gathered (<http://www.stat.fi>). This limitation is addressed later in the text where it is shown that the bias caused by the unequal representation of Finnish genders on the results of this study can be assumed to be minor. The relatively low response rate (17%) of Finnish students is probably due to a global trend of rising survey nonresponse rates (Massey & Tourangeau, 2013), and is another potential source of bias in this study. However, the results of Fosnacht, Sarraf, Howe, and Peck (2017) suggest that when at least 1000 students from each institution are invited to participate in a study, even a response rate of 5% should provide reliable information. Therefore, the Finnish response rate (17% with 17 922 students sampled at nine universities) can be

regarded as adequate for further analyses. The use of both online surveys and paper surveys in data collection should not have a major effect on the results, since earlier research has shown that collecting data in these two different modes only has a minor effect on survey responses (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003).

Data was collected using a questionnaire compiled by Kuittinen and Meriläinen (2011; Meriläinen, 2014; Meriläinen & Kuittinen, 2014). The questionnaire was translated into Chinese using English as an intermediate language, with semantic and syntactic adaptations made when necessary. The similarity of meanings was double-checked by the members of the Finnish-Chinese research team. This study focuses on three items measuring students' perceptions of **suitable workload** and 27 items measuring students' perceptions of their teaching-learning environment (items presented in Table 3 in Result section). The scales of the items range from *completely disagree* to *strongly agree* (1-5).

2.2 Statistical Analyses

This study is cross-cultural in nature, and therefore the level of structural equivalence of the research instrument between the cultural groups had to be examined first. In addition, when the levels of the perceived workload between the Finnish and Chinese students were compared, the impact that the different response styles of Finnish and Chinese students has on the results needed to be taken into account, since the differences between response styles can have a remarkable influence on the results in cross-cultural studies (He & van de Vijver, 2012).

The data of this study placed restrictions on statistical analyses, since the assumptions of standard statistical techniques were not fully met. This has been taken into account when choosing the methods, and robust versions of statistical analyses were employed when appropriate. Exploratory methods were used to examine the level of the structural equivalence of the research instrument between the Finnish and Chinese groups, since an explorative approach is recommended for cross-cultural studies in cases like the present, where the underlying dimensions are unclear (Fischer & Fontaine, 2011; He & van de Vijver, 2012; van de Vijver & Leung, 1997). Statistical analyses were undertaken using R version 3.2.2 (software package WRS downloaded 09/2015 from <http://dornsife.usc.edu/labs/rwilcox/software/>) and IBM SPSS statistics Version 23.

The investigation of the level of equivalence in the internal structures of the research instrument between the Finnish and the Chinese groups was initiated by determining the number of components to extract from the data of the Finnish group, which was assigned as the reference group. This was done by applying the commonly-used eigenvalue-greater-than-one -rule and by using Velicer's Minimum Average Partial Test (MAP) and Parallel analysis (computer programs published by O'Connor, 2000). The same number of components was then applied also for the Chinese group (using the procedure described by Fischer & Fontaine, 2011). The principal component analyses, which Fischer and Fontaine (2011) advocate for explorative purposes in cross-cultural studies, were then performed separately for the Finnish and Chinese students' data. However, since principal axis factoring is a method often used in studies of students' perceptions of TLE, it was also employed to verify the factor structures in both cultural groups. Since the components correlated in both groups, oblimin oblique rotation was used in both analyses. Finally, to be able to objectively evaluate the component similarity between the Finnish and the Chinese groups, a target rotation and a computation of the Tucker's phi coefficient were performed on the results of the analyses (computer program published by van de Vijver & Leung, 1997; also Fischer & Fontaine, 2011).

Once the level of equivalence of the research instrument was confirmed to be adequate, the differences in the mean levels of **suitable workload** between the student groups were investigated. First, the gender differences in **suitable workload** were examined within the Finnish and Chinese cultures using Wilcox' robust one-way ANOVA with 20% trimmed means (Wilcox, 2005). Second, the cultural comparisons between the Finnish and Chinese students were conducted both before and after taking into account the effects that the different response style of Finnish and Chinese students have on the difference in the mean levels of **suitable workload**. With the effect of the response styles not taken into account, the difference in **suitable workload** was examined using both the standard version and Wilcox' robust version (with 20% trimmed means) of one-way ANOVA. Partial eta squared and Cohen's d were used as standard effect sizes. The robust effect size used, ξ , is related to Cohen's d , so that for normally distributed data with homogeneous variances the values $\xi = 0.15, 0.35, \text{ and } 0.50$ correspond to the values $d = 0.2, 0.5, \text{ and } 0.8$, respectively (Wilcox, 2012). The response styles of Finnish and Chinese students were taken into account by using the method introduced by Reynolds and Smith (2010), which was selected among the various methods developed for this purpose since Van Vaerenbergh and Thomas (2013) described it as easy to use in their exhaustive review. The method controls for four measures: *extreme responding*, *mid-point responding*, *index of dispersion*, and *rating to the right of the center* (see Table 2 for more detailed information), by entering them as covariates in ANCOVA. Following the Reynolds and Smith -procedure, the impact of *extreme responding* was controlled for at the first stage

and the impact of all four response styles were controlled for at the second stage. Both the standard version and Wilcox' robust version of ANCOVA were used in parallel because the robust function *ancova* can only be applied with one covariate (Wilcox, 2005), and therefore the robust versions of statistical analyses could not be used all the way through the Reynolds and Smith -method. Hence, the effects of all four response styles were controlled for only within standard ANCOVA. The robust ANCOVA compares 20% trimmed means at five design points where the regression slopes of the covariate are the same, which ensures that the relationship between the outcome and covariate is approximately the same in both cultural groups (Field, Miles, & Field, 2012; Wilcox, 2005).

Table 2. Response style measures.

Extreme responding	% of responses that lie at the endpoints of the scale.
Midpoint responding	% of responses that fall in the middle category of the scale, i.e., 3 in the scale 1-5 (68 items), 3 and 4 in 1-6 (10 items), and 4 in 1-7 (4 items).
Rating to the right of center	% of responses that are strictly to the right of the middle category of the scale, i.e., 4 and 5 in the scale 1-5 (68 items), 5 and 6 in 1-6 (10 items), and 5-7 in 1-7(4 items). Positively and negatively worded items were computed separately and their average was used in order for the two types of items to have equal weight.
Index of dispersion	The index of dispersion D is defined as
	$D = \frac{C}{C-1} \left(1 - \sum_{i=1}^C r_i^2\right)$
	where
	C = the number of categories in the scale
	r_i = the ratio of responses that fall into the i:th category

Note: The measures were calculated for the Finnish and Chinese students separately using the unrecoded responses to all items of the questionnaire (82 items: 68(72) positively worded and 14(10) negatively worded items in the Finnish(Chinese) version).

Finally, moderated multiple regression (MMR) was used to examine the relationships between the six TLE variables and the variable **suitable workload**. Since teaching and learning environments of the Finnish and Chinese students are different, whereby the TLE variables are not directly comparable between the Finnish and the Chinese groups, the MMR analyses were conducted separately for the two cultural groups. In order to assess the gender differences within the cultures, the moderating role of gender on the relationships between TLE variables and the variable **suitable workload** was examined by constructing interaction terms between each of the TLE variables and the gender variable. Thus, a moderator (gender) variable and predictor (TLE) variables were entered into the analyses at Step 1 and the interactions were added at Step 2. The predictors were centered to reduce the possibility of multicollinearity. The moderator was coded, with 0 representing males and 1 representing females. Before conducting the MMR analyses the homogeneity-of-error-variances -assumption was tested across the moderator subgroups. This was done with computer program ALTMMR (Aguinis, 2004; Aguinis, Petersen, & Pierce, 1999). When the assumption was found to be violated, James's second-order approximation (U) and Alexander's normalized *t* approximation (A) were used in addition to MMR's *F*-test as is recommended by Aguinis (2004; Aguinis et al., 1999). The significance of the MMR was verified with bootstrapped (2000 replications) confidence intervals and standard errors. Since the MMR analyses were conducted separately for the Finnish and the Chinese groups, differences in response styles did not need to be taken into account. Also, for this reason, the numerical values of the results obtained from the analyses of the Finnish and the Chinese data cannot be compared to each other.

A number of statistical significance tests were conducted in this study, which increases the risk of Type I errors. The alpha levels, however, were not adjusted as such adjustments can be regarded as unjustified (see the reasoning by O'Keefe, 2003).

3. Results

3.1 Level of Equivalence of the Instrument

The criteria of the eigenvalue-greater-than-one rule indicated that seven components should be extracted from the Finnish data, whereas MAP and the Parallel analysis suggested that six components should be retained. Due to this inconsistency, a decision was made to proceed with both six and seven extracted components in the investigations. The principal component analyses with seven extracted components yielded a component structure which matches the factor structure of Kuittinen and Meriläinen (2011), whereas six extracted components generated a structure where **teaching quality** and **encouraging teachers** coalesce into one component, with the remaining components staying the same. Component similarity between the Finnish and Chinese groups was higher when seven components were extracted from the Finnish and the Chinese data, with the Tucker's phi coefficients varying from 0.90 to 0.96, which all are values regarded as adequate component congruence between cultural groups (Fisher & Fontaine, 2011; He & van de Vijver, 2012). With six extracted components the Tucker's phi coefficients vary between 0.50 and 0.97, so that the lowest values are less than 0.85, which indicates incongruence between the cultural groups (Fisher & Fontaine, 2011). Therefore a decision was made to use seven extracted components in the rest of this study. Principal axis factoring (factor loadings not reported here) produced similar factor structures when seven factors were extracted, with Tucker's phi coefficients varying between the acceptable 0.87 and 0.95. The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.916 for the Finnish group and 0.923 for the Chinese group, and the Bartlett's test of sphericity was significant ($p < 0.001$) in both cases. The Cronbach's alphas of the seven components vary from 0.69 to 0.89 for the Finnish group, and from 0.55 to 0.86 for the Chinese group. These values can be regarded acceptable (see Bowling, 2014). The component loadings and Cronbach's alphas of both groups and the Tucker's phi coefficient values are shown in Table 3.

3.2 Students' Perceptions of Suitable Workload

Statistically significant gender difference was found within the Chinese group ($Test=19.55, p < 0.001$), so that the trimmed mean level of perceived **suitable workload** was higher among male than female students, although the effect size was small ($\xi=0.133$). In contrast, no significant difference was found between the Finnish female and male students, which means that the unequal representation of the Finnish genders does not affect the results obtained when levels of perceived **suitable workload** are compared between the Finnish and the Chinese groups. The results of this cross-cultural comparison are reported both before and after controlling for the response styles in Table 4.

Before controlling for the response styles the results of standard and robust ANOVA indicate a statistically significant difference between the mean scores of Finnish and Chinese students, so that the mean level of **suitable workload** is higher in the Finnish group. The effect size is similar with both the standard and the robust method, being between small and medium. Controlling for the effect of *extreme responding* with standard ANCOVA has practically no effect on the difference between the mean scores of Finnish and Chinese students. After controlling for the effects of all four response styles, the difference in the mean levels increases and the effect size lies between medium and large. With robust ANCOVA the difference in the trimmed mean levels between the Finnish and Chinese students is statistically significant at all design points except for the first one (around 0.0). This suggests that there is no statistically significant difference in the trimmed mean levels between those Finnish and Chinese students who do not use the end points of the scale in their responses. The trimmed mean difference increases when the location of the design points shifted from 0 toward 1. This indicates that among those students who use the endpoints of the scales only seldom the difference between Finnish and Chinese groups is smaller than among those students who often use the end points of the scales. This illustrates that the Finnish students tend more often to use the high endpoints and the Chinese students tend more often to use the low endpoints when ranking their perception of **suitable workload**. The results of the robust version of ANCOVA vary in the neighborhood of the result obtained with the standard version, and thus discretion should be exercised when interpreting the results obtained here using standard ANCOVA. To summarize, irrespectively of whether the response styles are accounted for or not, these results indicate that Chinese students experience a heavier workload than Finnish students do.

Table 3. Results of principal component analyses of TLE

Item	Culture	Cr.α	Component loadings						
			1	2	3	4	5	6	7
Helpful evaluation	Finnish	.820							
	Chinese	.760							
●Evaluation of courses in major subject has supported my learning process	Finnish							.851	
	Chinese							.730	
●Evaluation of courses in major subject has supported me as a learner and enhanced my personal development	Finnish							.852	
	Chinese							.723	
●Evaluation of courses in major subject has corresponded well with my own perceptions of my achievement level	Finnish							.669	
	Chinese							.704	
Suitable workload	Finnish	.768							
	Chinese	.622							
●Workload in major subject as a whole is too much compared to study points (R)	Finnish								.870
	Chinese								.876
●Major studies' workload is at times unbearable (R)	Finnish								.759
	Chinese								.889
●Major studies' workload is reasonable and appropriate with the number of credits	Finnish								.794
	Chinese					.333			.367
Student relations	Finnish	.839							
	Chinese	.546							
●I have received encouragement, support and help from other students	Finnish							.890	
	Chinese							.734	
●I feel I belong to a group of students at my field	Finnish							.842	
	Chinese							.509	
●Discussions with other students has helped my learning	Finnish							.835	
	Chinese							.676	
●At my field it is difficult to get in touch with other students (R)	Finnish							.689	
	Chinese							.430	
Encouraging teachers	Finnish	.843							
	Chinese	.825							
●Teachers are usually sympathetic and kind to students	Finnish							.772	
	Chinese							.721	
●Teachers are supportive to students	Finnish						.422	.639	
	Chinese							.754	
●Teachers are patient and explain again difficult things if asked	Finnish					.393		.615	
	Chinese							.746	
●Teachers will appreciate the students' views and opinions on issues being studied	Finnish					.349		.626	
	Chinese							.711	
●Teachers are fair	Finnish							.702	
	Chinese							.599	.303

Table 3. continues

Item	Culture	Cr.α	Component loadings							
			1	2	3	4	5	6	7	
Teaching quality	Finnish	.887								
	Chinese	.862								
●Teachers have good teaching skills	Finnish	.708	.364							
	Chinese	.594								
●Teachers are interested in teaching, which is mediated to students	Finnish	.641	.446							
	Chinese	.522	.338	.305						
●Teachers are inspiring and engaging	Finnish	.534	.506							
	Chinese	.664								
●Teachers use a wide range of teaching methods	Finnish	.761								
	Chinese	.728								
●Teachers develop their teaching according to course feedback	Finnish	.600	.328							
	Chinese	.714								
●Teachers are interested in collecting teaching-related feedback	Finnish	.557	.418							
	Chinese	.710								
●Teachers seek to activate students' abilities and skills for active use	Finnish	.672								
	Chinese	.663								
●Teachers are able to adapt their teaching to match the level of students'	Finnish	.612								
	Chinese	.622								
Teaching for deep understanding	Finnish	.748								
	Chinese	.722								
●Teaching in my main subject emphasizes development and deepening of personal understanding in things studied	Finnish					.714				
	Chinese					.688				
●Teaching of my main subject inspire re-evaluate my understanding on the subject	Finnish					.748				
	Chinese					.696				
●Teaching of my main subject enhance learning, encouraging the student to form a personal understanding of issues taught	Finnish					.759				
	Chinese					.716				
●Teaching of my main subject has developed my skills as problem-solving	Finnish					.527				
	Chinese					.581				
Research-based teaching	Finnish	.686								
	Chinese	.763								
●Based on teaching of main subject I know what kind of research is done at our department	Finnish									.829
	Chinese									.782
●The teaching is clearly connected with the research activities of my department	Finnish									.770
	Chinese									.751
●It is easy for me to participate in the department's research activities within the framework of my studies	Finnish									.678
	Chinese									.777
Tucker's phi coefficient			.96	.93	.91	.94	.96	.93	.90	

Note: Loadings greater than or equal to .400 are shown in boldface and loadings smaller than .300 are excluded from the table.

3.3 Relationships between Perceived TLE and Perceived Suitable Workload

In the Finnish group the results obtained from ALTMMR showed that the homogeneity-of-error-variances -assumption is satisfied for all the moderator-based subgroups. The MMR analysis showed that the variables in the regression equation at Step 1 account for 11.7% of the variation in the dependent variable **suitable workload** ($\Delta F=57.20, p<.001$), with four out of six independent TLE variables, i.e., **encouraging teachers**, **helpful evaluation**, **teaching for deep understanding**, and **research-based teaching**, being significantly and positively related to **suitable workload**. No significant interactions were found in the Finnish group at Step 2 ($\Delta R^2=.002, \Delta F=.98, p=.436$).

In the Chinese group the MMR analysis showed that three out of six TLE variables, i.e., **helpful evaluation**, **student relations**, and **encouraging teachers**, are significantly and positively related to **suitable workload**, and the regression model explains 8.1% of the variation in **suitable workload** at Step 1 ($\Delta F=28.36, p<.001$). The results of the ALTMMR indicate that the error variances are heterogeneous in the moderator-based subgroups in those cases where the moderating effects of gender on the relationship between **helpful evaluation** and **suitable workload** and on the relationship between **student relations** and **suitable workload** were investigated. This means that the homogeneity-of-error-variances -assumption is violated in these cases, and therefore James's second-order approximation and Alexander's normalized t approximation were employed in addition to MMR's F -test. The result of the F -test showed statistically significant ($p=.031$) moderating effect of gender on the relationship between **student relations** and **suitable workload** at Step 2. In contrast, the results of James's test and Alexander's test showed no significant moderating effect. Since entering the interaction terms at Step 2 did not significantly increase the amount of variance explained ($\Delta R^2=.005, \Delta F=1.97, p=.067$), it can be stated, taking all these results together, that most likely there is no moderating effect of gender on the relationship between **student relations** and **suitable workload**. Finally, the results obtained from F -test, James's test and Alexander's test agreed in that the relationship between **helpful evaluation** and **suitable workload** is not moderated by gender.

Table 4. Differences in the mean levels of Suitable workload before and after taking response styles into account.

	Suitable workload				VARS ^a		
	MD ^b	Test	p	Effect size	X	n1	n2
Observed scores							
Standard ANOVA	.88	$F=131.72$	<.001	.024(η_p^2)			
				.324(d)			
Robust ANOVA	.99	$Test=125.69$	=.001	.234(ξ)			
Extreme responding taken into account							
Standard ANCOVA ^c	.85	$F=117.94$	<.001	.022(η_p^2)			
Robust ANCOVA	.22		=.058		0	833	1198
	.55		<.001		.13	1954	1782
	.81		<.001		.22	2077	1208
	1.43		<.001		.33	1520	731
	5.46		<.001		.82	35	40
All response styles taken into account							
Standard ANCOVA ^d	2.17	$F=616.79$	<.001	.104(η_p^2)			

a. VARS: variables of robust ANCOVA of Suitable workload. X: design points; n1, n2: number of cases at design points.

b. MD: mean difference. Values are positive when the Finnish mean value is greater than the Chinese mean value.

c. Covariate *extreme responding* was insignificantly related to Suitable workload.

d. Covariate *index of dispersion* was insignificantly related to Suitable workload.

The results of the MMR analyses are summarized in Tables 5 and 6 for the Finnish and the Chinese groups, respectively. An examination of the tolerance and the variance inflation factor statistics confirmed that multicollinearity was not a problem in the regression analyses. It can be concluded that the relationships between students' perceptions of teaching and learning environment and their perception of **suitable workload** are significant but relatively weak in both the Finnish and the Chinese groups. The relationships between the TLE variables and the **suitable workload** variable are not moderated by gender in either group.

Table 5. Moderated multiple regression analysis of the Finnish group. Confidence intervals (in parenthesis) and standard errors based on 2000 bootstrap samples.

Steps Variables	Suitable workload					U ^b	A ^c
	ΔR^2	B	SEB	β			
Step 1	.117***						
Gender		.03 (-.24, .28)	.14	.00 ^{//}			
Helpful evaluation		.14 (.08, .19)	.03	.11***			
Teaching for deep understanding		.12 (.07, .16)	.02	.12***			
Student relations		-.02 (-.05, .01)	.02	-.03 ^{//}			
Teaching quality		-.02 (-.05, .01)	.01	-.04 ^{//}			
Research-based teaching		.07 (.02, .11)	.02	.06**			
Encouraging teachers		.19 (.14, .23)	.02	.22***			
Step 2 ^a	.002 ^{//}						

* $p < .05$, ** $p < .01$, *** $p < .001$, ^{//} $p > .05$

a. No interactions were significant at this step.

b. James's second-order approximation

c. Alexander's normalized *t* approximation

Table 6. Moderated multiple regression analysis of the Chinese group. Confidence intervals (in parenthesis) and standard errors based on 2000 bootstrap samples.

Steps Variables	Suitable workload					
	ΔR^2	<i>B</i>	<i>SEB</i>	β	U ^b	A ^c
Step 1	.081***					
Gender		-.24 (-.42, -.06)	.10	-.05*		
Helpful evaluation		.10 (.04, .16)	.03	.10**		
Teaching for deep understanding		.01 (-.03, .06)	.02	.02 ^{//}		
Student relations		.14 (.09, .18)	.02	.15***		
Teaching quality		-.02 (-.04, .01)	.01	-.04 ^{//}		
Research-based teaching		-.04 (-.09, .01)	.02	-.04 ^{//}		
Encouraging teachers		.09 (.05, .13)	.02	.13***		
Step 2 ^a	.005 ^{//}					
Gender x Student relations		-.10 (-.20, -.01)	.05	-.08*	1.28 ^{//}	1.27 ^{//}

* $p < .05$, ** $p < .01$, *** $p < .001$, ^{//} $p > .05$

a. Only interactions with significant *F*-tests are included in the table.

b. James's second-order approximation

c. Alexander's normalized *t* approximation

4. Discussion

The students' perceptions of their workload and their teaching and learning environment were found to be related in both the Finnish and the Chinese groups, although the results imply that also other aspects of students' lives, not investigated in this study, clearly have an influence on students' perceptions of their workload. This is in line with the results obtained from Hong Kong by Kember and Leung (2006), and it supports their view that it may be possible to influence the students' perception of workload by improving the quality of their teaching and learning environments. The results show that evaluation that supports the students' learning as well as teacher encouragement reduce the perceived workload among both Finnish and Chinese students. Since these results were obtained when studying two educational systems with rather different cultural backgrounds, they can be assumed to be generalizable to university students in other countries as well. On the other hand, **teaching quality** was found not to be significantly related to **suitable workload** in either the Finnish or the Chinese group, which is in contrast to the research assumptions. This indicates that factors such as the use of a wide variety of teaching methods or the teachers' interest in students' feedback do not appreciably affect students' perceptions of their workload.

Notable differences between the cultural groups are that **student relations** was found to be related to **suitable workload** for Chinese students but not for Finnish students, whereas **teaching for deep understanding** and **research-based teaching** are related to **suitable workload** for Finnish students but not for Chinese students. Hence, in the first case, the research assumption was supported only for the Chinese group and in the latter cases only for the Finnish group.

One explanation for the observed positive relation between **student relations** and **suitable workload** for Chinese students may be that Chinese students associate good **student relations** more tightly to studying, and hence working together with peers or getting help from peers might indeed reduce the students' perceived workload (Kember, 2004), whereas in the Finnish group good **student relations** might be more closely linked to students' leisure activities and hence have less or no effect on the perceived workload. Another explanation could be the level of the perceived

workload of the Chinese students, which in this study was found to be higher than in the Finnish group. This could lead to a situation where only those Chinese students who perceive a suitable workload are able to frequently spend time with their peers, whereas those Chinese students who perceive a heavy workload can spend less time with their peers, which might have a negative effect on their relationships with other students (Kember, 2004). Finnish students, in contrast, may generally perceive a moderate enough workload for it not to affect their relationships with other students.

The explanations for the observed cultural difference regarding the relation between **teaching for deep understanding** and **suitable workload** and between **research-based teaching** and **suitable workload** may be rather similar. A lighter perceived workload may leave Finnish students more time and energy to actively participate in learning processes and hence to favor teaching, which aims to deepen students' understanding of the subject and integrates research in teaching. In the Chinese group, in contrast, an attempt to cope with a heavier perceived workload may push students to seek shortcuts and apply studying methods associated with a surface approach to learning (Kember, 2004), which may also reciprocally increase the perceived workload (Kember & Leung, 1998). A traditional, teacher-centered approach with transmission of knowledge is still a very common and a substantial element of teaching in China (Zhu, 2010; Yin, Wang, & Han, 2016), and, according to Yin et al. (2016), Chinese students have therefore become accustomed to and learned to prefer being passive recipients of traditional teaching. Consequently, it is possible that Chinese students have a tendency not to prefer teaching that aims for deep understanding and problem-solving skills regardless of whether they perceive their workload as heavy or not, which might explain why **teaching for deep understanding** is not significantly related to **suitable workload** in the Chinese group. The same explanation could be applied for the non-significant relation between **research-based teaching** and **suitable workload**, since increasing of the amount of integration of research with teaching would require the students to alter their behavior from being passive recipients of knowledge to taking active roles in constructing knowledge in collaboration with their teachers (Griffits, 2004).

In this study the question of causality in the studied relationships between the perceptions of the TLE elements and perceived workload is not easy to resolve, as the level of perceived workload may reciprocally affect the students' perception of their teaching-learning environment. This makes interpreting the results challenging. Nevertheless, a practical implication of these findings is that for students' to perceive a suitable workload, it is important that teachers in higher education are steered toward building encouraging relationships with their students and to use assessments, which support students' learning processes. The latter may be easier to achieve when new modes of assessment, such as portfolios, simulations, or self and peer assessment are used instead of the traditional multiple-choice and essay examinations (Struyven, Dochy, & Janssens, 2005), and positive bonds between teachers and students may form more easily when the course format is seminar, workshop or tutorial instead of traditional lecturing (Hagenauer & Volet, 2014). In addition, Chinese students may especially benefit from opportunities to study together with their peers, whereas Finnish students' perception of suitable workload is positively linked to teaching methods that enhance the students' deep understanding, problem-solving skills and research activity, and hence such methods should be employed in Finland.

5. Conclusion

In the present study it was investigated how Finnish and Chinese students' perceptions of workload are related to their perceptions of their local teaching and learning environments and the levels of perceived workload between the cultural groups were compared. In both groups the relationship between perceived TLE and workload was found to be statistically significant but relatively weak. Chinese students were found to experience heavier workload than Finnish students do. In China female students reported a slightly heavier perceived workload than did male students, whereas in Finland no gender difference was found.

Studying the relationships between perceived TLE and workload is of interest since an excessive workload can have a severe negative effect on the students' well-being and study success, which makes recognition of the means to reduce students' perceived workload valuable. A large number of university students from Finland and from mainland China were involved in this study. The comparison of Finnish and Chinese students is interesting, since Finland and China have different educational backgrounds, but both the Finnish and the Chinese educational systems have fared well in cross-national comparisons. In this study the statistical methods employed in the analyses were carefully chosen and robust statistical methods were used when the assumptions of the standard versions were not met. Because of the cross-cultural nature of this study, the level of equivalence of the measurement instrument between the cultural groups was examined and the effect of different response styles on the results was investigated. These are crucial aspects of cross-cultural studies, but especially the latter aspect is often neglected in research

papers. Therefore, it is important to continue to develop and use methods to minimize the effect of cultural differences in order to obtain unbiased results.

The present study has some limitations. The response rate of the Finnish group was relatively low (17%), which could cause nonresponse bias. However, earlier research has shown that in studies where the sampling frames are large, reliable information can be obtained even with a response rate of 5% (Fosnacht et al., 2017). Therefore the Finnish response rate can be taken as adequate for this study. Translating the questionnaire from Finnish to Chinese using English as an intermediate language may have caused bias in the data. Since Finnish and Chinese are linguistically and culturally rather different, a fair amount of cross-cultural adaptation was required. The process of translation and adaptation led four negative worded items in the Finnish questionnaire to be reversed to positive worded ones in the Chinese questionnaire. However, since these four items were not included in the items studied in this work, the effect of these reversals, which is expected to be minor, is restricted to the controlling of response bias. Furthermore, the averaging of positively and negatively worded items separately for both cultures presumably reduced the effect. Genders were unequally represented in the Finnish data. This limitation was, however, taken into account when conducting the analyses, and hence the bias caused by this can be regarded as minor. In addition, the Finnish and the Chinese groups differed somewhat in their ages and years of studies, which may also introduce some bias to the results of this study.

References

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.
- Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods*, 2(4), 315-339. <https://doi.org/10.1177/109442819924001>
- Bachman, L., & Bachman, C. (2006). Student perceptions of academic workload in architectural education. *Journal of Architectural and Planning Research*, 23(4), 271-304. <http://www.jstor.org/stable/43030781>
- Bowling, A. (2014). *Research methods in health: investigating health and health services*. 4th edition. Berkshire: Open University Press.
- Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College students' responses to web and paper surveys: does mode matter? *Research in Higher Education*, 44(1), 1-19. <https://doi.org/10.1023/A:1021363527731>
- Davies, J., Weko, T., Kim, L., & Thulstrup, E. (2009). *OECD reviews of tertiary education: Finland*. Paris: OECD.
- Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. Thousand Oaks, CA: Sage Publications, Ltd.
- Fisher, R., & Fontaine, J. R. (2011). Methods for investigating structural equivalence. In D. Matsumoto, & F. van de Vijver (Eds.), *Cross-cultural research methods in psychology* (pp. 179-215). New York, NY: Cambridge University Press.
- Fosnacht, K., Sarraf, S., Howe, E., & Peck, L. K. (2017). How important are high response rates for college surveys? *The Review of Higher Education*, 40(2), 245-265. <https://doi.org/10.1353/rhe.2017.0003>
- Gallagher, M., Hasan, A., Canning, M., Newby, H., Saner-Yiu, L., & Whitman, I. (2009). *OECD reviews of tertiary education: China*. Paris: OECD.
- Griffits, R. (2004). Knowledge production and the research-teaching nexus: the case of the built environment disciplines. *Studies in Higher Education*, 29(5), 709-726. <https://doi.org/10.1080/0307507042000287212>
- Gu, M. (2006). An analysis of the impact of traditional Chinese culture on Chinese education. *Frontiers of Education in China*, 1(2), 169-190. <https://doi.org/10.1007/s11516-006-0001-8>
- Hagenauer, G., & Volet, S. E. (2014). Teacher-student relationship at university: an important yet under-researched field. *Oxford Review of Education*, 40(3), 370-388. <https://doi.org/10.1080/03054985.2014.921613>
- He, J., & van de Vijver, F. (2012). Bias and equivalence in cross-cultural research. *Online readings in Psychology and Culture*, 2, 1-19. <https://doi.org/10.9707/2307-0919.1111>
- Jacobs, S. R., & Dodd, D. K. (2003). Student burnout as a function of personality, social support, and workload. *Journal of College Student Development*, 44(3), 291-303. <https://doi.org/10.1353/csd.2003.0028>
- Kember, D. (2004). Interpreting student workload and the factors which shape students' perceptions of their workload. *Studies in Higher Education*, 29(2), 165-184. <https://doi.org/10.1080/0307507042000190778>

- Kember, D., & Leung, D. Y.P. (1998). Influences upon students' perceptions of workload. *Educational Psychology, 18*(3), 293-307. <https://doi.org/10.1080/0144341980180303>
- Kember, D., & Leung, D. Y. P. (2006). Characterising a teaching and learning environment conducive to making demands on students while not making their workload excessive. *Studies in Higher Education, 31*(2), 185-198. <https://doi.org/10.1080/03075070600572074>
- Kuittinen, M., & Meriläinen, M. (2011). The effect of study-related burnout on student perceptions. *Journal of International Education in Business, 4*(1), 42-62. <https://doi.org/10.1108/18363261111170586>
- Massey, D. S., & Tourangeau, R. (2013). Where do we go from here? Nonresponse and social measurement. *The ANNALS of the American Academy of Political and Social Science, 645*(1), 222-236. <https://doi.org/10.1177/0002716212464191>
- Meriläinen, M. (2014). Factors affecting study-related burnout among Finnish university students: teaching-learning environment, achievement motivation and the meaning of life. *Quality in Higher Education, 20*(3), 309-329. <https://doi.org/10.1080/13538322.2014.978136>
- Meriläinen, M., & Kuittinen, M. (2014). The relation between Finnish university students' perceived level of study-related burnout, perceptions of the teaching-learning environment and perceived achievement motivation. *Pastoral Care in Education, 32*(3), 186-196. <https://doi.org/10.1080/02643944.2014.893009>
- O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, 32*(3), 396-402. <https://doi.org/10.3758/BF03200807>
- O'Keefe, D. (2003). Colloquy: should familywise alpha be adjusted? Against familywise alpha adjustment. *Human Communication Research, 29*(3), 431-447. <https://doi.org/10.1111/j.1468-2958.2003.tb00846.x>
- Parkinson, T. J., Gilling, M., & Suddaby, G. T. (2006). Workload, study methods, and motivation of students within BVSc program. *Journal of Veterinary Medical Education, 31*(2), 5-24. <https://doi.org/10.3138/jvme.33.2.253>
- Reynolds, N., & Smith, A. (2010). Assessing the impact of response styles on cross-cultural service quality evaluation: a simplified approach to eliminating the problem. *Journal of Service Research, 13*(2), 230-243. <https://doi.org/10.1177/1094670509360408>
- Shi, X., Xue, Z., & Zhang, H. (2015). A study on the research-oriented teaching courses reform in Chinese colleges and universities. *International Journal of Information and Education Technology, 5*(4), 260-264. <https://doi.org/10.7763/IJiet.2015.V5.513>
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment and Evaluation in Higher Education, 30*, 325-341. <https://doi.org/10.1080/02602930500099102>
- van de Vijver, F., & Leung, K. (1997). *Methods and data analysis for cross-cultural research*. Thousand Oaks, CA: Sage Publications, Inc.
- Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: a literature review of antecedents, consequences, and remedies. *Internal Journal of Public Opinion Research, 25*(2), 195-217. <https://doi.org/10.1093/ijpor/eds021>
- Watkins, D. (2003). Teacher thinking and practice from a Chinese cultural perspective: Lessons for East and West. In F. Salili, & R. Hoosain (Eds.), *Teaching, learning and motivation in a multicultural context* (pp. 243-258). Greenwich, CT: Information Age Publishing.
- Wilcox, R. (2005). *Introduction to robust estimation and hypothesis testing*. (2nd ed.). San Diego, CA: Academic Press.
- Wilcox, R. (2012). *Introduction to robust estimation and hypothesis testing*. (3rd ed.). Oxford, UK: Academic Press, Elsevier.
- Yin, H., Wang, W., & Han, J. (2016). Chinese undergraduates' perceptions of teaching quality and the effects on approaches to studying and course satisfaction. *Higher Education, 71*(39), 39-57. <https://doi.org/10.1007/s10734-015-9887-5>
- Zhu, C. (2010). Teacher roles and adoption of educational technology in the Chinese context. *Journal for Educational Research Online, 2*(2), 72-86.