

Development and evaluation of a Chinese version of the Questionnaire on Teacher Interaction (QTI)

Xiaojing Sun¹  · Tim Mainhard¹ · Theo Wubbels¹

Received: 23 January 2017 / Accepted: 3 July 2017 / Published online: 25 July 2017
© The Author(s) 2017. This article is an open access publication

Abstract Teacher–student interpersonal relationships play an important role in education. The Questionnaire on Teacher Interaction (QTI) was designed to measure students’ interpersonal perceptions of their teachers. There are two Chinese versions of the QTI for student use, and that inherited the weaknesses of the previous English versions, such as items that focus on class behaviour, instead of teacher behaviour, or that include conditionals and negations. The present study aimed to develop an improved Chinese version of the QTI which is conceptually parallel with the original QTI and with the use of optimal item wording. The process contained several steps, including expert panels and student and teacher interviews. New items were also created based on the Chinese secondary classroom context, rather than only using translations of English items. The final version of the Chinese version of the QTI presented in this paper was evaluated with a sample of 2000 students from 4 secondary schools in mainland China who rated a total of 80 teachers. The resulting version of the Chinese QTI had adequate validity and reliability, and it distinguished clearly between teachers. The predictive validity was supported by the relation between the students’ perceptions of their teachers and their academic emotions in class. Although further improvement of the instrument is recommended, it can be used to study interpersonal teacher behaviour in China and to help to improve Chinese teachers’ teaching practices.

Keywords Chinese secondary education · Interpersonal theory · Questionnaire on Teacher Interaction · Teacher–student relationships

✉ Xiaojing Sun
x.sun@uu.nl

¹ Department of Education, Utrecht University, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands

Introduction

Teacher–student relationships have been studied extensively in education (e.g. Maulana et al. 2012; Telli et al. 2007; Passini et al. 2015; Wubbels et al. 2014). The importance of interpersonal relationships in education has been appreciated for years because positive relationships between teachers and students contribute to student learning and wellbeing (Den Brok et al. 2004, 2005; Goh and Fraser 2000). To conceptualise teacher–student interpersonal relationships, Wubbels et al. (1985) adapted Leary’s (1957) interpersonal circle to the educational context into the Model for Interpersonal Teacher Behaviour (Wubbels et al. 1985) which, more recently and in line with research in interpersonal psychology, is also referred to as the *Teacher Interpersonal Circle* or the IPC-T (Pennings et al. 2014; Pennings and Mainhard 2016). The IPC-T is a circumplex model representing prototypical teacher behaviours. Based on this model, the *Questionnaire on Teacher Interaction* (QTI) was created to measure teacher–student interpersonal relationships by capturing students’ interpersonal perceptions of their teachers.

The QTI was originally developed in Dutch (Wubbels et al. 1985), and has been adapted to and translated into several other languages, including American English (Wubbels and Levy 1991), Australian English (Wubbels 1993), Turkish (Telli et al. 2007) and Indonesian (Maulana et al. 2012) (see Wubbels et al. 2014 for an overview). However, some QTI adaptations have involved straightforward, literal translations, thus heightening the risk of misunderstanding caused by variation in the interpretation of seemingly similar interpersonal meanings of words in different languages (Wubbels et al. 2012). According to Wubbels et al. (2012), a further step in the adaptation of the QTI is crafting conceptually parallel versions, rather than direct translations, by considering language and cultural embeddedness. This means that, besides high alpha reliabilities, it is also important for the adapted versions to have their items structured into a pattern that represents the circumplex nature of the model, and that the items represent equivalent positions on the interpersonal circle across versions.

Among the adaptations of the QTI are also Chinese versions. A simplified Chinese version was developed and applied in south-west China to measure the perceptions of Chinese students (Wei et al. 2009). A traditional Chinese version was created later in Hong Kong (Sivan et al. 2014). Earlier, a version specifically adapted to measure teachers’ self-perceptions was published in a Chinese journal (Xin and Lin 2000). Considering that the educational system and cultural context in China are very different from that in the Western countries, the goal of the present study was to further improve the current Chinese versions of the QTI in terms of reliability and validity, by crafting a Chinese version of the QTI which is conceptually parallel to the original QTI.

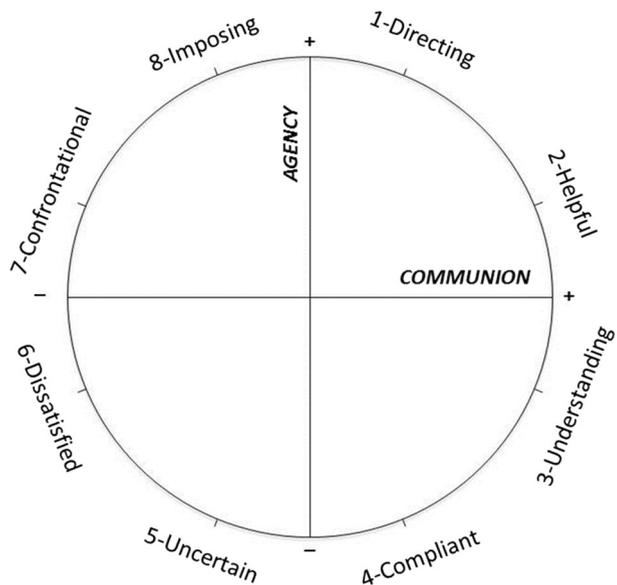
Teacher Interpersonal Circle (IPC-T)

The interpersonal theory was applied to conceptualise interaction and interpersonal perceptions in terms of the meta-labels *agency* and *communion*, which are combined in an interpersonal circumplex (IPC; Horowitz and Strack 2010). The *agency* (vertical) dimension refers to interpersonal influence and control, describing the level to which someone strives for dominance and control, and ranging from yielding to influencing. The *communion* (horizontal) dimension refers to interpersonal proximity, affiliation or connection, describing the degree of emotional togetherness someone conveys and ranging from

keeping separate to connecting with others. The IPC is a weighted combination of levels of both dimensions, reflecting all of the possible combinations of agency and communion (Wiggins 1991). In the educational context, the IPC for the teacher (IPC-T) is the latest version of the Model for Interpersonal Teacher Behaviour (Wubbels et al. 1985). As Wubbels et al. (2012) describe, during the last decades, a change in labels for these dimensions has occurred (Brekelmans et al. 2011; Wubbels et al. 1985, 2014), which reflects a process of compliance with the use of labels in interpersonal psychology (Horowitz and Strack 2010). At first, Influence and Proximity were used, later also control and affiliation were used and, finally, agency and communion were used as labels of the underlying dimensions of the IPC-T and thus the QTI. Note that this is merely a question of labels and not the essence of the underlying construct which remained the same. The IPC-T describes a teacher's general behavioural tendencies in interpersonal terms, and it can be used to describe students' interpersonal perceptions of how a teacher generally behaves in class (Wubbels et al. 2014). The IPC-T is divided into eight octants, describing eight prototypical types of teacher interpersonal behaviour as 1-Directing, 2-Helpful, 3-Understanding, 4-Compliant, 5-Uncertain, 6-Dissatisfied, 7-Confrontational and 8-Imposing (Mainhard 2015; Wubbels et al. 2014; see Fig. 1, numbers refer to octants in the model). Just as was the case with the underlying dimensions, also the labels for these eight specific positions on the IPC-T have changed somewhat over time with the aim of improving the clarity of what specific underlying blend of agency and communion is being referred to (Wubbels et al. 2014). Thus the original labels (Leadership, Helpful, Understanding, Student Freedom, Uncertain, Dissatisfied, Admonishing and Strict, in Wubbels et al. 1985) conceptually were meant to convey the same interpersonal meaning of teacher behaviour in class as the labels used more recently.

On the IPC-T, neighbouring octants are highly positively correlated (e.g. 8-Imposing and 1-Directing overlap substantially in their interpersonal meaning), whereas octants opposite to each other are highly negatively correlated (e.g. 1-Directing and 4-Uncertain). Theoretically and empirically, the two interpersonal dimensions are uncorrelated in

Fig. 1 Teacher Interpersonal Circle



heterogeneous samples, and the octants are equally distributed (i.e. equidistant), occupying specific positions on the circle. Each octant reflects a specific blend of the independent dimensions—agency and communion. Knowing the level of agency that a teacher conveys in class does not allow for inferring how communion is enacted, and vice versa. For example, teacher behaviour representing moderately high agency might either be confrontational behaviour (when combined with relative low communion) or helpful behaviour (when combined with relative high communion).

Questionnaire on Teacher Interaction (QTI)

The QTI was originally designed in Dutch and consisted of 77 items (Wubbels et al. 1985) measuring student perceptions (and teachers' self-perceptions) of the level of agency and communion that teachers convey in class. The items of the QTI are divided into eight scales corresponding with the eight octants of the IPC-T. Because each item corresponds to one of the octants, unlike in questionnaires with simplex structures, it loads on both underlying dimensions instead of only one. For example, the item "this teacher changes his/her mind in response to student feedback", belonging to the 4-Compliant scale, reflects both rather-low levels of agency and moderate levels of communion, whereas the item "this teacher gets angry quickly", belonging to the 7-Confrontational scale, reflects low teacher communion combined with moderately-high agency. Items are rated on a 5-point Likert-type scale, bounded by 'Never' and 'Always'.

The first translation of the QTI was an American English version, consisting of 64 items after adding, deleting and adjusting items based on several rounds of testing (Wubbels and Levy 1991). This first translation was a conceptually parallel version, with the items representing equivalent positions on the interpersonal circle as the original Dutch version. This version was initially also applied in Australia and was then revised into a 48-item selection for the Australian context (Wubbels 1993). This Australian version of the QTI was then translated into several languages and applied in various countries such as Singapore (Goh and Fraser 1998), Brunei (Scott and Fisher 2004), Turkey (Telli et al. 2007), Indonesia (Maulana et al. 2012) and Italy (Passini et al. 2015). Although these later translations could have had the goal of achieving to the original QTI parallel measures, most of them used translation procedures only.

Regarding conceptually parallel measures, Wubbels et al. (2012) noted that the original Dutch, and therefore the American and Australian versions and later translations, had some conceptual weaknesses. Most translators were not aware of these weaknesses which were published only in Dutch (Créton and Wubbels 1984). For instance, the correlations between octants (i.e. the position of octants on the circle) deviated from the position of octants on the theoretical circumplex in that some scales showed higher or lower correlations than a model with equidistant octant scores would exhibit. Another problem was related to the wording of items. According to Wubbels et al. (2012), the items should describe general, unconditional situations rather than specific instances of classroom situations (e.g. 'this teacher is uncertain'—5-Uncertain), focus on the teacher rather than students (e.g. 'this teacher has a sense of humour'—2-Helpful), concentrate on interpersonal processes rather than more didactic issues (e.g. 'this teacher is strict'—8-Imposing) and avoid using negative forms. However, for example, the Australian 48-item version contained 5 items using conditional formulations (e.g. 'if we don't agree with this teacher, we can talk about it'), 6 items describing student behaviour instead of teacher behaviour (e.g. 'we can decide some things in this teacher's class'), 3 items focussing on pedagogical or didactical rather than interpersonal issues (e.g. 'this teacher is severe when marking

papers') and 6 items used negative formulations (e.g. 'This teacher is not sure what to do when we fool around'). Versions adapted from the Australian adaptation have risked repeating these problems.

Furthermore, Wubbels et al. (2012) claimed that the goal of making an adaptation is to produce a conceptually parallel instrument with not only high alpha reliabilities for each scale, but also a pattern of scale correlations that represents the circumplex nature of the IPC-T. Therefore, the items need to be structured to measure each blend of agency and communion on the circle equally across all versions. Making parallel items by only translation and back-translation procedures might not be adequate to represent parallel circumplex structures across versions, because participants from different cultural contexts could interpret seemingly parallel translations as differing in the specific blend of agency and communion (i.e. the specific position of an item on the interpersonal circle). As noted by Telli et al. (2007), if only translation and back-translation procedures are applied, scale reliability might appear good, but it is questionable whether such a translation would be conceptually comparable to the original Dutch and English versions. Hence, when adapting the QTI to another language, it is essential to adapt to the local context by considering the position of the items on the circle, and thus to create a version conceptually parallel with the IPC-T.

With the essential goal of creating a questionnaire including items that are able to measure perceptions representing the eight octants of the IPC-T validly, conceptual parallel adaptations have been made in Turkey (Telli et al. 2007) and Indonesia (Maulana et al. 2012). For both versions, interviews with local teachers and students were performed, and different meanings of items of the original American English version as compared to the Turkish and Indonesian context were identified. New items were created based on interviews and some items were moved to other scales. About 70% of the original American English items were directly translated and used in the Indonesian version and about 40% were used in the Turkish version. More recently, a 24-item selection of the Dutch version was developed from the original 77 items after several rounds of item selecting process. This latest 24-item version has increasingly been used in most Dutch studies (e.g. Mainhard 2015; Pennings et al. 2014). These 24 Dutch items were selected and reworded based on the following criteria: describing general teacher behaviour, focusing on interpersonal process rather than pedagogical issues, concentrating on the teacher rather than the behaviour of students, and avoiding using negative forms. Thus, in terms of wording a more homogeneous set of items appeared. Also, the crafting of the Chinese version of the QTI described in the current paper was informed by these criteria.

Applications of the QTI in China

Until now, three Chinese versions of the QTI have been published, all of them using the 48-item Australian version as their starting point for translation and adaptation. Xin and Lin (2000) translated the Australian QTI into Mandarin and modified it into a teacher version with the specific aim of measuring teachers' self-perceptions only. A student version published by Wei et al. (2009) was applied in classrooms in the south-west part of China where English as a foreign language was taught. This was the first time that the QTI was applied in the Chinese context by accepted translation procedures with students assessing the behaviour of their teachers. Their sample consisted of 160 grade 8 students from four secondary education classrooms. For their version, single-level confirmatory factor analysis using MPlus (Muthén and Muthén 1999) was conducted to test the model fit at the scale level. A model that allowed scales to shift freely over the circle showed a rather

satisfactory fit whereas, for a stricter model applying the theoretical, equidistant positions of octant scores on the circumplex, the model fit was less satisfactory. For example, compared to the theoretical model, the 4-Compliant scale had a higher factor loading on Agency and the 5-Uncertain scale had a higher factor loading on the Communion dimension than expected (i.e. these scales were shifted over the IPC-T). Additionally, the internal reliabilities of some of the scales were not satisfactory and the authors concluded that additional improvement was needed. For example, more qualitative data sources were suggested to improve the quality of the questionnaire, such as interviews with students and teachers for generating new items and in order to assess whether translated items reflect actual teacher behaviour in the Chinese classroom context.

A different Chinese QTI adaptation translated from the 48-item Australian version was developed and tested by Sivan and Chan (2013), based on a convenience sample consisting of 612 grade 9 students from 16 classrooms in six secondary schools in Hong Kong. The validity was tested by calculating inter-scale correlations at both the individual student level and the class level. Although the correlations generally reflected the circumplex nature of the IPC-T, some problems were observed. For instance, the 1-Directing scale was positively correlated with all seven of the other scales at the individual level, indicating that a problem with measuring more submissive teacher behaviours existed. In sum, in terms of the circumplex structure, the correlations between scales indicated a deviant spacing and ordering of octants. More recently, Sivan et al. (2014) improved the instrument with a sample of 739 grade 5 and 6 primary school students in Hong Kong. Principal component analysis with oblique rotation and maximum likelihood confirmatory factor analysis were performed on two subsamples separately and one-, three-, and eight-factor models were examined. Although the eight-factor model showed a satisfactory fit, it indicated many correlated factors rather than two orthogonal factors (i.e. agency and communion) underlying the IPC-T. Both student versions of the Chinese QTI adaptations inherited the item-formulation problems of the original Dutch and Australian version.

The current study's goal was to develop an improved Chinese version, a conceptual parallel version which well represents the circumplex nature of the IPC-T and with the use of optimal item wording. The process included: first, following the procedure of item formulation and selection as applied to the improvement of the most recent Dutch version (e.g. focusing on general, in-class behaviours of teachers and avoiding the use of negative forms); second, ensuring face validity of items in the Chinese classroom setting, by conducting interviews with students and teachers to test whether the translated items, octant and dimension labels were able to represent the intended combination of agency and communion in the actual classroom context; and, finally, applying a stricter, confirmatory test of the circular structure of items and scales to improve the validity of the questionnaire.

Therefore, in the current study, a Chinese version of the QTI was developed based on the previous versions by reformulating existing items but also by crafting new items. This version was tested with a large data set. The research question was: "To what extent is the newly-developed Chinese version of the QTI a reliable and valid instrument for measuring students' interpersonal perceptions of their teachers in Chinese secondary classrooms?"

Methods

Item crafting and face validity

As a first step, informed by the previous Chinese versions and on the basis of the Australian version and the Dutch 24-item selection, and following the item crafting criteria as described above, 37 Chinese items were crafted by the first author. Two other Chinese educational researchers were involved in ensuring the face validity for the Chinese context.

To test whether the wording of items represented well the intended blend of agency and communion in the classroom context, and to receive additional practical suggestions for item wording, semi-structured interviews, based on the existing 37 items, were conducted with 10 teachers and 10 students from two regular high schools and two vocational secondary schools in Jining city, Shandong Province, which is located in the eastern part of China. Nine female students and one male student participated in the interviews, with their ages ranging from 15 to 18 years. The participating teachers included nine female teachers and one male teacher, and their teaching experience varied from 3 to 30 years. All interviewees participated on a voluntary basis and the data were treated anonymously. The interviews were audio recorded with the oral agreement of every interviewee. To assess how items were interpreted, the participants were asked to think-out loud while completing the questionnaire. Teachers were asked to think about how they perceived themselves in class, while students were asked to think about how they perceived their favourite and least favourite teachers while discussing the items offered and while thinking of new or alternative item formulations. For example, favourable teacher behaviours described by students were as follows:

This teacher is very responsible in class. Every time when we had questions, she would explain clearly and figure out why we had these questions.

She treats all students equally. No matter whether a student has good grades or bad grades, she shows no partiality.

He is very strict on our study, but he cares about us in daily life and he is like a friend to us.

Examples of unfavourable teacher behaviours described by students were as follows:

She is always late for class and never apologises for it. Also, she doesn't care whether we understand what she teaches in class, which is not very responsible.

She says offensive words to students. She wants us to respect her, but she doesn't respect us.

She punishes us by making us clean the classroom for a week if we are late for school, or if we are found to use our cell phone in class.

Based on these interviews, most of the items were considered understandable, but also several problems appeared. For example, according to the interviewees, some 4-Compliant and 5-Uncertain items were problematic because the behaviours described in these items almost never occurred in their classrooms. This was in line with the low reliabilities of these two scales in earlier Chinese versions of the QTI (Sivan et al. 2014; Wei et al. 2009). To solve this problem, the participants were asked to think of situations in which a teacher might be compliant or uncertain in the participants' eyes, and several items were reworded.

Starting with this round of interviews and with the goal of creating a larger item base for quantitative testing and selection of items, an item pool consisting of 80 items was created.

Main test

Sample and procedure

The 80 Chinese QTI items were administered to 2000 grade 7–9 students from 40 classrooms in 4 public junior-secondary schools in Weihai city, Shandong Province in May 2015. First, approval was provided by the principals to conduct the survey. Further, it was made clear to teachers and students that participation was voluntary and that the data would be treated anonymously. An administrative teacher went into each classroom to distribute paper questionnaires and answer sheets to students during self-study classes.

In order to assess the predictive validity of the QTI, additionally to the QTI items, eight academic emotion items were administered to measure emotions experienced by students during class. Academic emotions are the emotions experienced by students in academic settings on a daily basis, and these emotions are connected with teacher behaviour in classrooms (Pekrun et al. 2006). In the present study, two academic emotions were used: enjoyment and anxiety. Four items were selected for each emotion from a Chinese version of the Academic Emotions Questionnaire (AEQ) (Pekrun et al. 2005). Items were answered on a 5-point Likert-type scale. Finally, 1995 usable questionnaires were collected. The students on average were 13.49 years old, ranging from 11 to 17 years. Of the students, 50.6% were females, 1427 students (71.5%) were in grade 7, 231 (11.6%) were in grade 8, 196 (9.8%) were in grade 9, and 141 students (7.1%) failed to report the grade level. As is common in China, each class contained about 50 students; approximately half of the students in each class (22–26 students) completed the questionnaire for a specific teacher, resulting in ratings of 80 teachers. Each school received a short report summarising the results at a general level.

Item selection

To develop a shorter, less time-consuming, reliable and valid instrument, item selection was performed from the 80 items in the pool. For this selection, the sample was split into two subsamples by sorting the data by teacher, labelling the even- and odd-numbered questionnaires for each teacher to samples A and B, respectively. Item selection, based on internal validity analysis and reliability analysis, was performed on sample A ($n = 1029$), and internal validity and reliability were then re-evaluated with sample B ($n = 964$).

To select items, the first step was to assess the quality of each item using item descriptive statistics, including the mean, range, standard deviation, skewness and kurtosis. Intra-class correlations (ICC) were calculated in order to examine the consensus between students and thus whether an item could distinguish between teachers adequately. Items with a relatively low ICC (less than 0.10) or extreme descriptive coefficients (e.g. skewness greater than 2.0, kurtosis greater than 3.0) were nominated for exclusion. For the QTI, items are viewed as repeated measures of an octant; therefore, internal reliabilities and item-rest correlations for each octant were also considered.

CircE (Grassi et al. 2010) was used in the R statistical environment, version 3.2.2, to select items based on how the items and scales were projected on the IPC-T and to test the overall model fit. CircE was specifically developed to test the structural validity of items with an underlying circumplex structure. Model fit was checked with four goodness-of-fit

indices after each round of trimming: the standardised root mean square residual (SRMR); the root mean square error of approximation (RMSEA); the comparative fit index (CFI); and the Tucker-Lewis index (TLI). Because the RMSEA levies a harsher penalty for complexity in relatively small models containing only a few variables (Kline 2011), we used the RMSEA at the item level but emphasised it less when evaluating models based on scales.

Both unconstrained and stricter constrained models were tested. The unconstrained model represented a free circumplex, hypothesising two independent dimensions yet allowing scales to shift over the circle. In the stricter model, items were constrained to be situated within a specific octant, and the octants were set to be spaced equally around the circle (Grassi et al. 2010). After each round of item trimming, the unconstrained and constrained models were also tested at the scale/octant level for the purpose of checking the order and spacing of the scale scores. Additionally, the internal reliabilities of the scales involved were checked. In some cases, items with less than favourable descriptive coefficients or relatively low ICC values had to be retained to maintain acceptable validity and scale reliability. For instance, the items “this teacher let students boss him/her around” (skewness = 3.38) and “this teacher tolerates a lot of student behaviour” (ICC = 0.08) were included in the final item selection.

Some items were projected in a different octant than originally intended, with their loadings indicating a more favourable fit in a different octant. In this case, when face validity allowed, these items were moved from the original to the other octant. For instance, the original 7-Confrontational item “this teacher is easily offended” was projected in the 6-Dissatisfied octant on the circle and was better represented as measuring dissatisfied behaviour. So we moved it into the 6-Dissatisfied scale in the questionnaire. The item “this teacher can take a joke” originally belonged to the 2-Helpful scale in the Dutch, American and Australian versions, whereas in the current Chinese sample it loaded on the 4-Compliant octant. Items were removed when they loaded in a different octant than intended but face validity didn’t allow moving them. In the end, 12 items were moved between octants.

This process of trimming was repeated until no further improvement of the model with as few items as possible could be reached. Of the 40 items included in the final selection of the Chinese QTI, 13 were newly-created items (see “Appendix” section for a list of all included items). The validity and reliability of this selection of items were then re-evaluated on subsample B.

Finally, the scale and dimension scores were correlated with students’ academic emotions to evaluate the predictive validity of the current QTI version. The Cronbach alpha reliability for enjoyment and anxiety emotions were 0.86 and 0.57, respectively. As stated by Pekrun et al. (2006), enjoyment emotions tend to be experienced when an activity is valued as positive and warm, while anxiety is usually experienced when an activity is perceived to be negative and cold. Therefore, it could be anticipated that the Enjoyment scales would be correlated positively with the four QTI scales representing positive communion—1-Directing, 2-Helpful, 3-Understanding and 4-Compliant—and negatively with the other four scales representing negative communion, while the Anxiety scale was anticipated to be correlated with the QTI scales reversely. Both Enjoyment and Anxiety were expected to have stronger correlations with the octants representing the most and least communion, namely, 2-Helpful, 3-Understanding, 6-Dissatisfied and 7-Confrontational.

Results

The validity was tested with the CircE application in R in both unconstrained and constrained conditions at the item and scale levels for samples A and B. The values of the four fit indices of the Chinese QTI are listed in Table 1.

From Table 1, it can be seen that, in the unconstrained model for items in sample A, all the indices suggested adequate fit. When all of the items were constrained to fall within octants, only the TLI had a slightly lower value. Also for scale scores, except for the RMSEA, all indices showed supportive values for good model fit in both the unconstrained and constrained conditions. Note that we deemed the RMSEA less informative for models including scale scores (see Sect. [Item Selection](#)). Overall, the indices showed that the questionnaire adequately represented the IPC-T in sample A. As might be expected, because the model fit was maximized for sample A, the fit indices were somewhat less supportive in sample B. The model fit, however, was still deemed acceptable.

Figure 2 provides a visual representation of how the QTI items projected on the IPC-T in both the constrained and unconstrained situations. The items and scale scores situated on the left side of the IPC-T (representing negative communion) fell within their octants as anticipated. However, the spacing was not perfect. Further, a few items of the 1-Directing, 3-Understanding, 4-Compliant and 8-Imposing scales fell on the octant boundaries. At the scale level, the score for 3-Understanding also located on the upper border of the respective octant, indicating that it was correlated too strongly with the 2-Helpful scale. Apparently, the 3-Understanding items were slightly more agentic than they should be theoretically, but this is consistent with the versions in other countries, such as Singapore and Australia (Den Brok et al. 2006). Furthermore, the spacing between the 1-Directing and 8-Imposing scales was greater than expected. From Fig. 2, we can see that there was a lack of items representing sufficiently high agency situated in the upper middle space of the circle indicating, based on a theoretical basis, larger-than-desirable communion differences between these two scales.

Amongst the eight scales, 1-Directing, 4-Compliant, 5-Uncertain and 8-Imposing scale scores spaced slightly closer to the centre of the circle, indicating that, for this sample, items did reflect somewhat small variance. Amongst them, the 4-Compliant scale was located closest to the centre of the circle because of lower variances of its items on both agency and communion, demonstrating a problem similar to one in previous studies in

Table 1 Goodness-of-fit indices for CircE models of the Chinese QTI

Model	RMSEA	SRMR	TLI	CFI
<i>Sample A</i>				
Item (unconstrained)	0.04	0.06	0.90	0.91
Item (constrained)	0.05	0.08	0.89	0.90
Scale (unconstrained)	0.10	0.04	0.92	0.97
Scale (constrained)	0.10	0.04	0.92	0.97
<i>Sample B</i>				
Item (unconstrained)	0.05	0.06	0.89	0.90
Item (constrained)	0.05	0.09	0.88	0.89
Scale (unconstrained)	0.13	0.05	0.89	0.96
Scale (constrained)	0.14	0.08	0.85	0.95

Because the RMSEA levies a harsher penalty for complexity in relatively small models containing only a few variables (Kline 2011), the RMSEA should be less emphasised when evaluating the models based on scales

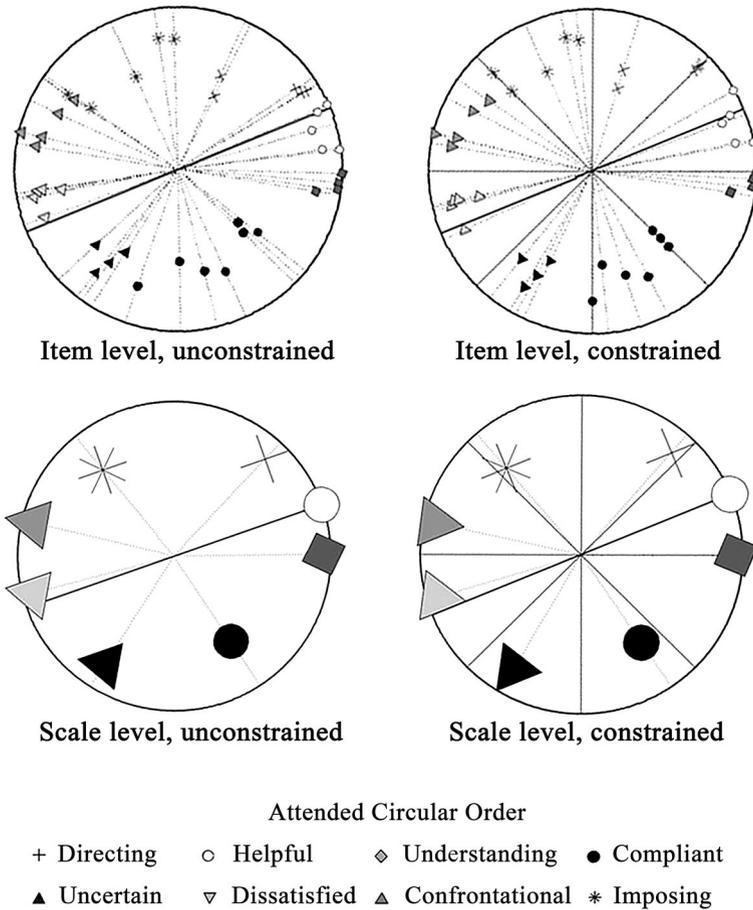


Fig. 2 Circular models of the 40-item Chinese QTI as displayed by CircE (Grassi et al. 2010) for sample A

China and other countries, such as Singapore, Brunei, Australia (Den Brok et al. 2006), Turkey (Telli et al. 2007), Indonesia (Maulana et al. 2012) and Italy (Passini et al. 2015).

Figure 2 shows the order and spacing of items and scales in a circumplex under two different conditions: constrained and unconstrained (see labels). The two models on the right side (the constrained models) contain the eight boundaries representing the theoretical borders of the octants. Cronbach alpha coefficients ranged between 0.60 and 0.84 for the eight scales, showing acceptable reliabilities for both sample A and sample B (for an overview, see Table 2). For sample A, the ICC at the octant level ranged from 0.13 to 0.31, indicating that the questionnaire could distinguish rather well between teachers. Besides the ICC, which refers to the average correlation between individual students' ratings of the same teacher, the ICC2 was also calculated, providing an estimate of the reliability of class-mean ratings (Lüdtke et al. 2009). The ICC2 indicated that the classroom aggregates of the scales were rather reliable (values greater than 0.70 indicate sufficient reliability, Lüdtke et al. 2009), ranging from 0.79 to 0.92 for an average class size of $n = 25$, and between 0.74 and 0.82 for $n = 10$ except for the 4-Compliant scale (0.60). The reliabilities of both the agency and communion dimensions were adequate, and the correlation between

Table 2 Mean, standard deviation, ICC and reliability of the Chinese QTI scales and dimensions

Scale	Number of items	M	SD	ICC	ICC2 (<i>n</i> = 10)	ICC2 (<i>n</i> = 25)	Cronbach's alpha	
							Sample A	Sample B
1-Directing	5	0.72	0.21	0.23	0.75	0.88	0.68	0.64
2-Helpful	5	0.84	0.21	0.31	0.82	0.92	0.84	0.84
3-Understanding	4	0.71	0.28	0.25	0.77	0.89	0.83	0.81
4-Compliant	7	0.31	0.17	0.13	0.60	0.79	0.61	0.60
5-Uncertain	4	0.10	0.16	0.25	0.77	0.89	0.61	0.60
6-Dissatisfied	5	0.11	0.18	0.22	0.74	0.88	0.78	0.81
7-Confrontational	5	0.21	0.23	0.26	0.78	0.90	0.80	0.81
8-Imposing	5	0.47	0.21	0.21	0.73	0.87	0.63	0.62
Agency	–	0.20	0.12	0.30	0.81	0.91	0.70	0.66
Communion	–	0.33	0.20	0.31	0.82	0.92	0.91	0.92

Cronbach's alpha was calculated at the student level. The mean scores of the dimensions were scaled to range between -1 and 1 . As is customary, dimensions were calculated by applying to each scale the theoretical weight ($a = 0.923880$, $b = 0.382683$) of the theoretically assumed circumplex: Agency = ($a*1$ -Directing) + ($b*2$ -Helpful) – ($b*3$ -Understanding) – ($a*4$ -Compliant) – ($a*5$ -Uncertain) – ($b*6$ -Dissatisfied) + ($b*7$ -Confrontational) + ($a*8$ -Imposing); Communion = ($b*1$ -Directing) + ($a*2$ -Helpful) + ($a*3$ -Understanding) + ($b*4$ -Compliant) – ($b*5$ -Uncertain) – ($a*6$ -Dissatisfied) – ($a*7$ -Confrontational) – ($b*8$ -Imposing) (see Wubbels and Brekelmans 2005)

the two dimensions was statistically nonsignificant with a value of 0.03 ($p = 0.28$). See Table 2 for an overview.

To check the predictive validity of the current QTI version, correlations between the QTI and AEQ scales were inspected. As anticipated, the 1-Directing, 2-Helpful, 3-Understanding and 4-Compliant scales were positively correlated with Enjoyment and negatively correlated with Anxiety, whereas the 5-Uncertain, 6-Dissatisfied, 7-Confrontational and 8-Imposing scales were correlated with the two emotions reversely. The scales representing the highest communion (2-Helpful, 3-Understanding) had the strongest positive correlations with Enjoyment and negative correlations with Anxiety, while the scales representing the lowest communion (6-Dissatisfied, 7-Confrontational) had the strongest negative correlations with Enjoyment and positive correlations with Anxiety. Hence, predictive validity results also supported the adequacy of the instrument (see Table 3).

Discussion

The purpose of this study was to craft an improved instrument to measure teacher agency and communion and their eight underlying related octants in the Chinese context. Overall, the reliability and validity of this 40-item Chinese QTI were supported by the data analyses. The strengths of this newly developed questionnaire are as follows. First, the following more-stringent criteria for item wording were followed from the very start of the development of the current version: describes general, unconditional situations rather than specific instances of classroom situations; focuses on the teacher rather than students; concentrates on interpersonal processes rather than more didactic issues; and avoids using negative formulations (Wubbels et al. 2012). Second, by performing interviews with

Table 3 Correlations of academic emotions scales with QTI scales and dimensions

Scale	Enjoyment	Anxiety
1-Directing	0.39*	-0.18*
2-Helpful	0.59*	-0.34*
3-Understanding	0.61*	-0.34*
4-Compliant	0.21*	-0.12*
5-Uncertain	-0.30*	0.22*
6-Dissatisfied	-0.42*	0.32*
7-Confrontational	-0.51*	0.42*
8-Imposing	-0.14*	0.21*
Agency	0.06	0.04
Communion	0.64*	-0.42*

* $p < 0.05$

students and teachers, the current version was explicitly grounded in the Chinese classroom context; new items were formulated based on these interviews. Third, the present study was the first to apply a strict and constrained model-fit testing approach that is specifically for assessing the circular fit of items and scales and suited to testing the specific assumptions of this specific questionnaire. Nevertheless, there is still room for improvement.

In the future, more work should be undertaken to improve the properties of some items. In general, the correlations (i.e. spacing) between the scales could still be improved, especially between the 8-Imposing and 1-Directing scales and between the 4-Compliant and 5-Uncertain scales. In terms of the underlying dimensions, it proved difficult to formulate a sufficient number of items reflecting very high/low agency with moderate communion (situating at the very top/bottom of the circle). Researchers in the future should attempt to find more agentic items for both 1-Directing and 8-Imposing, such as the good fitting items “this teacher is strict” and “this teacher’s standards are very high” in the current version, as well as more submissive items for both 4-Compliant and 5-Uncertain, such as the item “this teacher let students do what they want”. This challenge was also reflected in the relatively lower reliability of the agency dimension as compared with the communion dimension in the current version.

The 1-Directing, 4-Compliant, 5-Uncertain and 8-Imposing items showed relatively small variance in the current sample. A possible reason might be the involved teachers’ homogeneity in terms of agency in the current sample, which also could relate to the relatively high power-distance in the Chinese school culture (Hofstede et al. 2010). Also, as became evident in the interviews, almost all of the teachers seemed to show similarly high agency while describing their classroom behaviours, and several teachers mentioned that high agency was a property that a qualified teacher always had. It is possible that agency is a more important selection criterion for teachers in China than in Western countries. Thus, for further testing of the Chinese QTI, having a more heterogeneous sample in terms of teacher agency is important.

Special attention should be paid to the most problematic scale, namely, 4-Compliant. Already in the interviews, it became apparent that it was difficult for students and teachers to think of a compliant Chinese teacher because compliant teacher behaviours are rarely experienced in the classroom. In the end, the 4-Compliant scale had more items than other scales, but still had the lowest reliability. All seven items had a relatively low variance, and four of them were situated on the octant border. In addition, a few 4-Compliant items still contained wording problems regarding the item crafting criteria, because the most fitting

Chinese wording tended to describe what students did in such a teacher's class rather than describing the teacher. For example, the item "this teacher lets students get away with a lot" in Chinese translation says "in this teacher's class, students' mistakes can be let go". This translation was attributed to the same corresponding translation of the word "let" and "ask someone to do something" in the Chinese language, which is more active than the compliant behaviour the item aimed to describe. Future research should therefore pay additional attention to the formulation and selection of the 4-Compliant items.

As stated by Wubbels et al. (2012), for research purpose, 16 well-chosen items might already be sufficient to measure students' interpersonal perceptions of teacher agency and communion. From this point of view, the current Chinese version is still rather time-consuming. Future researchers might further attempt to make a more efficient measure with fewer items. Moreover, to further validate this instrument, studies could also be carried out on how students' perceptions measured with the questionnaire are aligned with actual teacher–student interactions in class (Pennings et al. 2014; Pennings and Mainhard 2016).

Notwithstanding the limitations and challenges discussed, we think that the current version of the questionnaire is already an improvement and suitable for obtaining students' perceptions of their teachers' agency and communion in Chinese secondary classrooms for research and to provide Chinese secondary-school teachers with feedback about their teaching. The approach applied in this study might also be informative for future adaptations of other instruments. Ultimately, after studies on relationships between student outcomes and their perceptions of teacher agency and communion in China have been conducted, it might be possible to design specific measures for creating positive teacher–student relationships with favourable student outcomes specifically for the Chinese context, thus providing some guidance to secondary-school educators for the improvement of teaching and learning in China.

Acknowledgements This study was supported by grants from the Chinese Scholarship Council.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Table 4.

Table 4 The 40 items of the Chinese QTI (with English translation)

Scale	Item (Chinese)	Item (literal English translation)
	这位老师	This teacher...
1-统领	...有威信。	...is prestigious
1-Directing (5 items)	...对课堂上的一切了如指掌。	...knows everything that goes on in the classroom
	...对学生的领导力强。	...shows good leadership with students
	...掌控学生课上的各种行为。	...controls students' behaviour in class
	...掌控学生何时能够讲话。	...controls when students can speak

Table 4 continued

Scale	Item (Chinese) 这位老师	Item (literal English translation) This teacher...
2-支持 2-Helpful (5 items)	...处事公正。 ...能抓住学生的注 意力。 ...令人敬重。 ...很可靠。 ...很好地掌控课 堂。	...is fair/impartial ...holds students' attention ...is respected ...is reliable ...manages class well <i>Note</i> This item is inspired by the original English item "This teacher is a good leader"
3-体谅 3-Understanding (4 items)	...幽默风趣。 ...上的课让人愉 快。 ...愿意倾听学生的 心声。 ...满面笑容。	...has a sense of humour ...class is pleasant ...listens to students ...has a smiling face
4-顺从 4-Compliant (7 items)	...顺应学生的要 求。 ...课上学生有很大 自由。 ...容忍学生的很多 行为。 ...可以和他/她开玩 笑。 ...根据学生的意见 改变主意。 ...课上学生的错误 可以被放过。 ...课上学生可以做 自己想做的事。	...is compliant to what students want ...students have a lot of freedom in class ...tolerates a lot of student behaviour ...can take jokes ...changes his/her mind in response to student feedback ...students' mistakes can be let go in class <i>Note</i> This item is inspired by the original English item "This teacher lets students get away with a lot" ...students can do what they want in class
5-犹疑 5-Uncertain (4 items)	...容易被学生捉 弄。 ...被学生牵着鼻子 走。 ...课上学生可以开 小差。 ...对纪律要求低。	...is easy to be made fool of by students ...let students boss him/her around ...students can skive in class <i>Note</i> This item is inspired by the original English item "This teacher lets students fool around in class" ...has low requirements on discipline <i>Note</i> This item is inspired by the original English item "This teacher's discipline is weak"
6-不满 6-Dissatisfied (5 items)	...说话不算话。 ...满腹牢骚。 ...推卸责任。 ...容易和学生起冲 突。 ...猜疑学生。	...break his/her words ... complains a lot ...passes the buck ...is easy to get conflicts with students <i>Note</i> This item is inspired by the original English item "This teacher is easily offended" ...is suspicious of students

Table 4 continued

Scale	Item (Chinese) 这位老师	Item (literal English translation) This teacher...
7-对抗	...容易发火。	...gets angry quickly
7-Confrontational (5 items)	...强行占用课余时间。	...forcibly occupies spare time
	...严厉处罚学生。	...punishes students severely
	...用惩罚来威胁学生。	...threatens students with punishment
	...令人害怕。	...is fearsome
8-强权	...很严格。	...is strict
8-Imposing (5 items)	...对学生要求非常高。	...standards are very high
	...要求学生服从他。	...requires students for obedience
	...强制课堂保持安静。	...imposes silence in class
	...强制学生按照其要求去做。	...forces students to do as what he/she says

The items above are ordered by the eight scales. Items should be randomised before using the questionnaire. The English translation is added for clarification, but *these are not the actual items used in the English versions of the QTI*

References

- Brekelmans, M., Mainhard, T., den Brok, P., & Wubbels, T. (2011). Teacher control and affiliation: Do students and teachers agree? *The Journal of Classroom Interaction*, *46*, 17–26.
- Créton, H. A., & Wubbels, T. (1984). *Discipline problems with beginning teachers*. Utrecht: WCC. (in **Dutch Ordeproblemen bij beginnende leraren**).
- Den Brok, P., Brekelmans, M., & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. *School Effectiveness and School Improvement*, *15*(3–4), 407–442.
- Den Brok, P., Fisher, D., Wubbels, T., Brekelmans, M., & Rickards, T. (2006). Secondary teachers' interpersonal behaviour in Singapore, Brunei and Australia: A cross-national comparison. *Asia Pacific Journal of Education*, *26*(1), 79–95.
- Den Brok, P., Levy, J., Brekelmans, M., & Wubbels, T. (2005). The effect of teacher interpersonal behaviour on students' subject-specific motivation. *The Journal of Classroom Interaction*, *40*(2), 20–33.
- Goh, S. C., & Fraser, B. J. (1998). Teacher interpersonal behaviour, classroom environment and student outcomes in primary mathematics in Singapore. *Learning Environments Research*, *1*(2), 199–229.
- Goh, S. C., & Fraser, B. J. (2000). Teacher interpersonal behaviour and elementary students' outcomes. *Journal of Research in Childhood Education*, *14*, 216–231.
- Grassi, M., Luccio, R., & Blas, L. D. (2010). CircE: An R implementation of Browne's circular stochastic process model. *Behaviour Research Methods*, *42*(1), 55–73.
- Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind*. New York: MacGraw-Hill.
- Horowitz, L. M., & Strack, S. (Eds.). (2010). *Handbook of interpersonal psychology: Theory, research, assessment, and therapeutic interventions*. New York: Wiley.
- Kline, R. B. (2011). *Principles and practice of structural equation modelling* (3rd ed.). New York: The Guilford Press.
- Leary, T. (1957). *An interpersonal diagnosis of personality*. New York: Ronald Press Company.
- Lüdtke, O., Robitzsch, A., Trautwein, U., & Kunter, M. (2009). Assessing the impact of learning environments: How to use student ratings of classroom or school characteristics in multilevel modeling. *Contemporary Educational Psychology*, *34*(2), 120–131.

- Mainhard, T. (2015). Liking a tough teacher: Interpersonal characteristics of teaching and students' achievement goals. *School Psychology International*, 36(6), 559–574.
- Maulana, R., Opendakker, M. C. J. L., Den Brok, P., & Bosker, R. J. (2012). Teacher–student interpersonal relationships in Indonesian lower secondary education: Teacher and student perceptions. *Learning Environment Research*, 15, 251–271.
- Muthén, B., & Muthén, L. (1999). *MPlus: Users guide* (2nd ed.). Los Angeles, CA: Muthén & Muthén.
- Passini, S., Molinari, L., & Speltini, G. (2015). A validation of the Questionnaire on Teacher Interaction in Italian secondary school students: The effect of positive relations on motivation and academic achievement. *Social Psychology of Education*, 18, 547–559.
- Pekrun, R., Elliot, A. J., & Maier, M. A. (2006). Achievement goals and discrete achievement emotions: A theoretical model and prospective test. *Journal of Educational Psychology*, 98(3), 583–597.
- Pekrun, R., Goetz, T., Frenzel, A. C., & Perry, R. P. (2005). *Academic Emotions Questionnaire (AEQ): User's manual—Chinese version*. Unpublished manual, University of Munich, Germany.
- Pennings, H. J. M., Brekelmans, M., Wubbels, T., Van der Want, A. C., Claessens, L. C. A., & Van Tartwijk, J. (2014). A nonlinear dynamical systems approach to real-time teacher behaviour: Differences between teachers. *Nonlinear Dynamics, Psychology, and Life Sciences*, 18(1), 23–45.
- Pennings, H. J. M., & Mainhard, M. T. (2016). Analyzing teacher–student interactions with state space grids. In M. Koopmans & D. Stamovlasis (Eds.), *Complex dynamical systems in education*. New York: Springer.
- Scott, R. H., & Fisher, D. L. (2004). Development, validation and application of a Malay translation of an elementary version of the Questionnaire on Teacher Interaction. *Research in Science Education*, 34, 174–194.
- Sivan, A., & Chan, D. W. K. (2013). Teacher interpersonal behaviour and secondary students' cognitive, affective and moral outcomes in Hong Kong. *Learning Environments Research*, 16, 23–36.
- Sivan, A., Chan, D. W. K., & Kwan, Y. W. (2014). Psychometric evaluation of the Chinese version on the Questionnaire on Teacher Interaction (C-QTI) in Hong Kong. *Psychological Reports: Measures & Statistics*, 114(3), 823–842.
- Telli, S., Den Brok, P., & Cakiroglu, J. (2007). Students' perception of science teachers' interpersonal behaviour in secondary schools: Development of a Turkish version of the Questionnaire on Teacher Interaction. *Learning Environments Research*, 10(2), 115–129.
- Wei, M., Den Brok, P., & Zhou, Y. (2009). Teacher interpersonal behaviour and student achievement in English as a foreign language classrooms in China. *Learning Environments Research*, 12, 157–174.
- Wiggins, J. S. (1991). Agency and communion as conceptual coordinates for the understanding and measurement of interpersonal behaviour. In W. W. Grove & D. Cicchetti (Eds.), *Thinking clearly about psychology: Vol. 2. Personality and psychotherapy* (pp. 89–113). Minneapolis, MN: University of Minnesota Press.
- Wubbels, T. (1993). *Teacher–student relationships in science and mathematics classes*. Perth: Curtin University of Technology.
- Wubbels, T., Brekelmans, M., Den Brok, P., Levy, J., Mainhard, T., & Van Tartwijk, J. (2012). Let's make things better: Developments in research on interpersonal relationships in education. In T. Wubbels, M. C. Opendakker, & P. Den Brok (Eds.), *Interpersonal relationships in education* (pp. 225–249). Rotterdam: Sense Publishers.
- Wubbels, T., Brekelmans, M., Den Brok, P., Wijsman, L., Mainhard, M., & Van Tarwijk, J. (2014). Teacher–student relationships and classroom management. In E. T. Emmer & E. Sabornie (Eds.), *Handbook of classroom management* (2nd ed., pp. 263–386). New York: Routledge.
- Wubbels, T., Créton, H. A., & Hooymayers, H. P. (1985). *Discipline problems of beginning teachers: Interactional behaviour mapped out*. Paper presented at the annual meeting of the American Educational Research Association, Chicago (ERIC Document Reproduction Service Np. Ed. 260040).
- Wubbels, T., & Levy, J. (1991). A comparison of interpersonal behaviour of Dutch and American teachers. *International Journal of Intercultural Relations*, 15, 1–18.
- Wubbels, T., & Brekelmans, M. (2005). Two decades of research on teacher–student relationships in class. *International Journal of Educational Research*, 43(1), 6–24.
- Xin, Z., & Lin, C. (2000). 教师互动问卷中文版的初步修订及应用 [Preliminary revision and application of the Questionnaire on Teacher Interaction]. *心理科学 [Psychological Science]*, 23(4), 404–407 (in Chinese).