

*International Journal of Learning, Teaching and Educational Research*  
Vol. 19, No. 11, pp. 109-126, November 2020  
<https://doi.org/10.26803/ijlter.19.11.7>

## Comparison of Students in Teacher Education from China and the USA: An Assessment of Dispositions

**Judy R. Wilkerson**

Florida Gulf Coast University, Ft. Myers, Florida  
<https://orcid.org/0000-0001-5484-6747>

**Lasonya L. Moore**

University of South Florida, St. Petersburg, Florida  
<https://orcid.org/0000-0001-5476-3511>

**W. Steve Lang**

University of South Florida, Florida  
<https://orcid.org/0000-0003-2208-6223>

**Jingshun Zhang**

Florida Gulf Coast University, Ft. Myers, Florida  
<https://orcid.org/0000-0002-1002-2743>

**Abstract.** This study explored whether differences in teacher candidate dispositions exist between Chinese and American students, while continuing validation of the updated *Beliefs About Teaching Scale* (BATS2). BATS2 incorporated the Rasch model of item response theory on Thurstone dichotomous items to measure commitment to the Interstate Teacher Assessment and Support Consortium Standards (InTASC) along the levels of the Bloom/Krathwohl affective taxonomy. This research is a unique combination of theories and practice – standards-based items, affective taxonomy, and modern measurement theory. Differential group function (DGF), applied in a mixed methods design, confirmed national differences, indicating differential commitment to standards and items. For standards and items that showed a difference in the two groups, literature and cultural context supporting those differences was identified to frame the qualitative portion of this study. For example, US teachers were more averse to assessment, clearly the result of the focus on standardized testing that is so resented in the US; the Chinese were less compelled to master content, which is less imperative in early childhood programs (the sample in this study). Results can be used in considering training needs and making instructional design more likely to be impactful for US institutions training Chinese natives and for Chinese institutions updating programs based on international input.

**Keywords:** accreditation; affective assessment; teacher education standards; dispositions; Rasch model

## 1. Introduction

When adults are asked who was their favorite teacher in school, most make that selection based on the teacher's dispositions (LaPaglia, 2020; Wasicsko, 2007) – empathy, enthusiasm, motivation, self-efficacy – to name a few. The importance of teacher dispositions has been recognized for decades; Wasicsko (2007) suggested that when an administrator hires a teacher with the correct dispositions, students learn and develop, parents are pleased, and district administrators are able to focus on the business of education (LaPaglia, 2020).

In the early 21<sup>st</sup> century, attention to the assessment of teacher dispositions was evolving. A number of assessments designed to measure preservice dispositions were developed (Edgington & Cox, 2015), typically using single instrument item types (Bair, 2017; Eberly, Rand & O'Connor, 2007; Wilkerson & Lang, 2007). Assessments such as the *Teacher Disposition Index* (TDI) developed by Schulte, Edick, Edwards and Mackiel (2004) and the *Eastern Teacher Dispositions Index* (ESTDI), developed by Singh and Stoloff (2008), are typical of these self-report surveys.

Lang and Wilkerson (2006) argued that single instruments and more systematic adherence to principles of sound measurement as well as national teaching standards were needed. Using the US national teaching standards (InTASC or Interstate Teaching and Assessment Support Consortium) Standards, developed by the Council of Chief State School Officers (CCSSO, 2013), they built a battery of five assessments of teacher dispositions, based on a five-step, standards-based assessment design process. Both the battery (Lang & Wilkerson, 2006) and the process (Wilkerson & Lang, 2007) bear the name, *Dispositions Assessments Aligned with Teacher Standards*, or DAATS.

Over the years, Wilkerson and Lang were joined by various researchers and universities in validating and using the instruments in the battery, most notably the *Beliefs About Teaching Scale*, or BATS, now in its second edition (see for example, Englehart et al., 20011; Lang & Wilkerson, 2006, 2008; Parfitt, et al, 2019; Wilkerson & Lang, 2006; Wilkerson, 2012). These studies examined the individual use of BATS as well as scaling using Rasch measurement to combine scores from multiple instruments in the battery.

The research teams also examined predictable differences across populations, including evidence that teacher dispositions were more consistent with the InTASC Standards in programs that specifically taught dispositions as opposed to programs that did not (Parfitt et al, 2019), regardless of degree level. In this latter study, the highest scoring respondents were masters' level students who were taught national educational leadership standards and not InTASC standards. These masters' level students "outperformed both undergraduate and doctoral level students. Authors concluded that teaching standards-based

dispositions (as opposed to not focussing on them) was what was important and not the specific set of standards applied.

No studies have been conducted, however, to examine whether differences in populations of different national backgrounds could be explained by national standards and culture. Both the US and China have national standards that incorporate dispositions, presenting an appropriate comparison group for the study presented herein. Therefore, this study continues research on measuring teacher dispositions (Lang et al., 2018; Lang & Wilkerson, 2006; Wilkerson, 2012; Wilkerson & Lang, 2007), adding an international cultural dimension.

## **2. Purpose and Significance**

The purpose of this study is to examine the response patterns of U.S. and Chinese nationals on a previously validated measure of teacher dispositions, designed for U.S. nationals but now tested with Chinese nationals. This study adds to the validation of the instrument; however, the primary purpose of this research is to determine if the quantitative patterns are explainable through a qualitative analysis of the standards themselves within national cultural context. Two universities, one in each country, serve as the sampling locations in this mixed methods study.

The findings are of new interest in exploring how the two cultures view their roles as teachers based on their respective national standards, which have marked similarities but also marked differences. It also continues to underscore the viability of measuring (not just assessing) teacher dispositions, providing additional evidence of the validity of the scores obtained from the BATS instrument.

## **3. Teacher Training and National Standards in the United States**

For nearly 60 years, the public education policy in the United States has been driven by federal legislation driving testing. The Elementary and Secondary Education Act (ESEA), enacted in 1965 and reauthorized periodically, has provided substantial funding for schools and focused on evaluating the progress of underserved students. Transitioning to the No Child Left Behind (NCLB) Act in 2002, this legislation was the genesis of formal educational evaluation with its focus on the promotion of high standards and increased accountability. The Every Student Succeeds Act (ESSA) followed in 2015, sharing NCLB goals but with more flexibility. Student learning has been centered on the Common Core Standards, developed by consortia of states, and used in the development of high stakes, standardized tests that determine promotion or failure of students in the public school system (Popham, 2020).

Teacher education programs are also held accountable for the application of national standards. To achieve national accreditation in the USA, Educator Preparation Programs (EPPs) must demonstrate graduates' competency in the Interstate New Teacher Assessment and Support Consortium (InTASC) Standards developed by the Council of Chief State School Officers (CCSSO, 2013). These Standards establish expected knowledge, skills, and critical dispositions for

beginning in-service teachers. Accreditation by the Council for Accreditation of Educator Preparation (CAEP, 2016), applied nationally and internationally, requires use of these Standards and dispositions for approved teacher education programs.

CCSSO defines dispositions as the “habits of professional action and moral commitments that underlie the performances and play a key role in how teachers do, in fact, act in practice” (CCSSO, 2013, p. 6). As a result, an essential responsibility of CAEP accredited teacher education programs has become the assessment of preservice teacher dispositions throughout the program (Almerico, Johnston, Henriott & Shapiro, 2011; Behrmann & Souvignier, 2013; Boonen, Van Damme & Onghena, 2014; Cox, Cheser, & Detwiler, 2015; Vaughn, 2012; Wilkerson & Lang, 2011).

There are multiple teacher training models in the US. This research is most closely aligned with the four-year university programs that are accredited. Some four-year programs are not accredited. There are also alternative means to teacher certification including course sequences offered at the universities and on-site training offered by the districts.

#### 4. Teacher Training and National Standards in China

The last 40 years of Chinese national curriculum reforms have targeted improved student academic performance through new classroom teaching models replacing or supplementing the didactic, teacher-centered approach with its strong focus on compliance (Han, 2012; Hu & Verdugo, 2015; Tian-Ping, 2012; 2011; Zhu & Han, 2006). Among the changes is the 2011 Ministry of Education of national teacher standards. These are remarkably similar to the InTASC Standards, providing for knowledge, skills, and dispositions. Han (2012) presented the Chinese dimensions and sub-categories, re-created in Table 1, followed by Table 2, listing the InTASC categories and standards.

**Table 1: Dimensions and subcategories of Chinese NCSTE  
(Ministry of Education)**

| <b>Dimensions of Curriculum Goals</b>       | <b>Subcategories of Each Dimension</b>   |
|---|--|
| 1: Educational beliefs and responsibilities | 1.1: Demonstrating scientific beliefs in children development and corresponding behaviors.<br>1.2: Demonstrating scientific beliefs in the profession of teachers and corresponding behaviors.<br>1.3: Demonstrating scientific beliefs in education and corresponding behaviors |
| 2: Educational knowledge and skills         | 2.1: Demonstrating knowledge and skills about understanding students<br>2.2: Demonstrating knowledge and skills for instruction students.<br>2.3: Demonstrating knowledge and skills for professional development.   |

|   |   |
|---|---|
| 3: Educational practice and experiences | 3.1: Developing experiences of observations in clinical fields.<br>3.2: Developing experiences of participating in educational practice.<br>3.3: Developing experiences of studying educational practice. |
|---|---|

**Table 2: InTASC Categories and Standards**

| InTASC Categories              | InTASC Standards  |
|--------------------------------|---|
| 1: The Learner and Learning    | 1. Learner Development<br>2. Learning Differences<br>3. Learning Environments     |
| 2: Instructional Practice      | 4. Assessment<br>5. Planning for Instruction<br>6. Instructional Strategies       |
| 3: Content Knowledge           | 7. Content Knowledge<br>8. Application of Content                                 |
| 4: Professional Responsibility | 9. Professional Learning and Ethical Practice<br>10. Leadership and Collaboration |

The Chinese dimensions correspond to dispositions (defined as beliefs), knowledge and skills, and practice and experiences. Each of the InTASC Standards is comprised of a one-page long set of attributes for performances, essential knowledge, and critical dispositions. While both countries focus on performance, knowledge and skill and dispositions, the order is noticeably different. Dispositions are first for the Chinese and last in the USA.

The first two Chinese dimensions are aligned well with three of the InTASC standards, although there is less focus on the second InTASC category related to content. The last Chinese category is well aligned with the American focus on clinical practice, CAEP Standard 2, although that is not specifically a part of the InTASC Standards. There is a strong focus on what is identified as professionalism and professional development in the American standards, similar to the third Chinese dimension. This reconceptualization of pedagogical thinking requires education programs that focus on educator dispositions through individual levels of cultural persistence (Gavin & Wang, 2016).

In China, there are multiple routes to teaching (Yu, 2013). Normal universities offer teacher education programs. Students studying in other multidisciplinary universities may also become teachers through the academic departments. Typically, the departments of education prepare teachers for early childhood education and primary school education, with the other departments responsible for higher grades. Different subject area teachers are used in all schools except at the primary level, where there is a single teacher for all subjects.

There are four teacher education models in China (Yu, 2013). The 2+2 model includes specialization courses in the first two years and pedagogy in the third and fourth years. The 2.5+1.5 model emphasizes subject theory for the first 2.5 years and pedagogy after. The 3+1 model has subject area learning for the first 3

years and pedagogy in the remaining time. There is also a fourth model, 4+2, at Beijing Normal University, with four years in subject and two in pedagogy, leading to a bachelors in a subject area and a master's in education. Most teacher preparation programs in the USA follow a 2+2 model, providing for general education during the first two years and teacher preparation in the last two years.

## 5. Methodology

This study used a mixed method design to determine if nationality made a difference in teacher education students' response patterns to a Thurstone (1928) survey. The quantitative portion used both the Rasch model of item response theory; the qualitative portion used a logical analysis that sought to explain or justify differences based on culture and national standards. The study was guided by the following questions:

1. What is the difference in the commitment of Chinese and American teacher education students to critical teacher dispositions, as defined by the InTASC Standards? (quantitative component)
  - a. What is the difference in commitment across all standards?
  - b. What is the difference in commitment for any of the ten standards?
  - c. What is the difference in commitment for any of the 50 specific items?
2. In instances where there are statistically significant differences, what logical explanations can be posited based on the national culture or standards? (qualitative component)

For research question #1, Rasch item response theory was applied to calculate scores, using Winsteps software (Linacre, 2018). The analysis included both Welch's *t*-test and differential group functioning (DGF) to determine group differences and statistical significance. With the focus in Rasch theory on both people and items (Rasch, 1960), the mathematical structure provided for calculation of the difficulty of the items and the ability (or commitment) of the people concurrently. This allowed for an analysis of responses at the item and standard level so that Chinese and US candidates could be compared on each question and each standard to determine if there was a statistically significant difference (contrast) in responses. For research question #2, a logical analysis was used to identify possible explanations for differences.

### 5.1 Instrument

The *Beliefs About Teaching Scale* (BATS) is a Thurstone agreement scale of 50 items, developed using the *Dispositions Assessments Aligned with Teacher Standards*, or DAATS model (Wilkerson & Lang, 2007). It is one of a series of five instruments, comprising the DAATS Battery (Wilkerson & Lang, 2006), assessing InTASC critical dispositions (2013).

Instruments comprising the DAATS battery are being revised and re-validated using the revised versions of the InTASC Standards from 2013. BATSv2 uses the new version of the Standards, but the structure theory for the five item types is the same as in the original DAATS. That application was described previously, along with discussions of analyses of different item types, evidence of validity,

and estimates of reliability (Lang & Wilkerson, 2008; Wilkerson, 2012). Another study (Englehart et al., 2012) provides sample items, descriptions of the five instruments, quantitative analysis, and qualitative discussions.

BATS is the easiest instrument to administer and score with dichotomous items. Technically, items are scored as right/wrong but are interpreted more legitimately as consistent/inconsistent with the InTASC critical dispositions. The remaining four instruments include a reflection on teaching experiences (*Experiential Teaching Questionnaire, ETQ*), a projective (*Situational Reflection Assessment, SRA*), an observation (*Candidate Belief Checklist, CBC*), and a focus group with children (*K-12 Impact Dispositions Scale, KIDS*). While BATS is easy to administer and score, it is also subject to faking, since students can provide the answer that they think is right, whether they believe it or not. The other instruments in the battery are increasingly difficult to fake and score.

The DAATS model calls for careful construction of items along a continuum of commitment based on the affective taxonomy (Krathwohl, Bloom & Masia, 1973). BATS version 2 (Forms A and B) was more systematic in its construction of items, retaining all that performed well statistically and adding new ones to fill coverage gaps. Evidence of validity and reliability has been presented previously for both versions of the scale. (See, for example, Wilkerson, 2012; Lang et al., 2014 and 2018).

## 5.2 Sample

The sample included 649 American and 84 Chinese teacher education students at varying levels of internship experience in teacher training programs. The Americans included attended two Florida universities, which are public universities in a state university system, and have multiple colleges including Education, and offer bachelors, masters, and doctoral degrees.

The Chinese students were enrolled in a provincial university similar in size to the two American universities. China has four levels of universities, and it is at the third level. It has one specialty with distinctive features at the national level, and a variety of other specialties with various levels of recognition, including a college of education that trains teachers for early childhood education. Other students planning to teach were enrolled in the subject area departments. The university has a strong research focus, with 60 research projects granted at the provincial level. Adhering to the policy of opening up to the outside world, the Chinese university has currently established international cooperation programs with 18 universities and research institutions from many other countries. This Chinese university uses the 3+1 model for teacher training, with the fourth year only dedicated to teacher preparation. The Chinese students completed BATS2 in May at the conclusion of their fourth academic year. The US students all completed BATS2 in September.

## 5.3 Data Analysis

The research utilizes a conjoint, IRT (item response theory) called the Rasch model (1960) that calibrates both person ability and item difficulty on an interval scale or “ruler”. The Rasch model estimates separation reliability, and can be used to

establish validity with standards-based assessments used with individual examinees. With Rasch modeling, the normality of the distribution is not assumed. While Rasch modeling most often is used to provide diagnostic information on how well test items measure affective values, and it is useful to analyze dichotomous, partial credit, and rating scale items together or separately. The Rasch model can also provide estimates of judge error in ratings (Eckes, 2015), which is of interest with the DAATS battery since four of the five instruments involve a rating.

The data were analysed using Winsteps, providing item calibrations, student measures, and difference by the independent variable of nationality. The first segment used differential group function (DGF) for the 10 item subgroups (InTASC Standards) and the individual items. DGF is a powerful technique for identifying latent classes among the persons (Linacre, 2018) and has been successfully used to detect country differences in large scale testing (Vanes, 2012).

The Real Item Separation Reliability of the analysis (N=733) =.99 while the Real Person Separation Reliability =.68. The result supports that the items created a defined variable and the scale discriminates adequately between the persons (Smith & Wind, 2018). Item Infit (.99) and Item Outfit (.93) was slightly less than the expected value of 1.0. Person Infit (.98) and Person Outfit (.93) were also slightly less than the expected value of 1.0. This may mean a less variability in the fit than expected by the model, and individual item fit is examined below.

## 6. Findings across All Standards

Research Question 1a asked if there was a statistically significant difference in commitment to the entire set of InTASC Standards, each Standard individually, and each item individually. Table 3 presents the output from Winsteps software, retyped and re-formatted. It includes the number of respondents, the mean Rasch measures (scores), standard deviation, median, model separation, model reliability for Chinese and US teacher candidates separately and combined.

**Table 3: Rasch Analysis of Chinese and American Students**

| Person Count | Mean Measure | Standard Deviation | Median | Model Separation | Model Reliability | Group    |
|--------------|--------------|--------------------|--------|------------------|-------------------|----------|
| 733          | 67.62        | 7.63               | 69     | 1.53             | .70               | All      |
| 649          | 68.39*       | 7.53               | 69     | 1.47             | .68               | American |
| 84           | 61.73*       | 5.62               | 62     | 1.12             | .56               | Chinese  |

*Note.* American-Chinese Mean Measure difference = 6.66, df=125, t=9.77. \*p<.000

The overall reliability is .70 (with reliabilities for US and China .68 and .56 respectively because of the sample characteristics where sample ability variance affects model statistics and wider ability range usually result in higher person reliability (Linacre, 2018). Chinese teacher candidates scored lower than American teacher candidates. A Welch's t-test, useful in situations with unequal variances in the groups, was used to test the hypothesis of equal means. The mean measure difference was 6.66 (df=125, t=9.77, p<0000), indicating that the difference between the two nationalities is statistically significant for the overall measure, including all ten Standards.



More persons, including some with no interest in teacher education, would provide for greater variability and higher person reliability. In fact, the entire sample is a relatively homogeneous group, so the scores of these teacher candidates are relatively close, with self-selected students generally interested in teaching. The item reliability is high at .98, indicating excellent variability (separation) between the items, with some items much more difficult than others.

In response to Research Question 2 for the overall set of standards, these results were as expected because of the shorter program in China, which targets predominantly early childhood teachers. Furthermore, US candidates are expected to demonstrate the InTASC Standards, while Chinese students use other similar but different teacher standards. Means were slightly more than one standard deviation apart, an understandable difference, given this context.

### 6.1 Findings for Individual Standards

Research Question #1b asked if there was a difference in commitment to each of the ten InTASC Standards. Differential group functioning analysis was performed using Winsteps. This procedure is used to check if items are working differently for two groups and, if so, where the most conspicuous differences are. The DGF contrast is the statistic of interest. It presents the number of logits of difference between the two groups. A logit (log odds unit) is the unit of measurement in Rasch, so from Table 1, the mean measure is 68.39 logits, and the total group standard deviation is 7.63 logits. Any contrast less than one-half of a standard deviation (less than 3.82 logits) was defined to be of minimal interest and not considered for further analysis. Omitted also are contrasts that are not statistically significant.

In Table 4, the DGF results are presented. Positive and negative DGF contrasts are assigned based on the order of entry into the software, indicating only the group for which the item was more difficult. For the population on the left (US), the value will be positive if the item(s) were more difficult for them. For the population on the right (Chinese), the value will be negative if the item(s) were more difficult for them. The positive or negative value has no numerical meaning. Again, a Welch t-test was conducted to check for statistical significance.

**Table 4: Rasch Analysis by InTASC Standards**

| Nation    | DGF Size | Nation       | DGF Size | DGF Contrast | Rasch-Welch t | DF  | Probability | InTASC Standard |
|-----------|----------|--------------|----------|--------------|---------------|-----|-------------|-----------------|
| US        | -.43     | <i>China</i> | 1.84     | -2.27        | -1.19         | 913 | .2330       | 1               |
| <i>US</i> | .00      | China        | -.76     | .76          | .46           | 675 | .6433       | 2               |
| US        | -.51     | <i>China</i> | 3.11     | -3.63        | -3.13         | 979 | .0018       | 3               |
| US        | -2.13    | <i>China</i> | 9.76     | -11.89       | -6.23         | 364 | .0000*      | 4               |
| <i>US</i> | .32      | China        | -2.12    | 2.44         | 2.09          | INF | .0370       | 5               |
| <i>US</i> | .72      | China        | -4.99    | 5.71         | 4.66          | 925 | .0000*      | 6               |
| US        | -1.15    | <i>China</i> | 6.89     | -8.04        | -5.33         | 490 | .0000*      | 7               |
| <i>US</i> | .27      | China        | -1.83    | 2.10         | 1.70          | 945 | .0902       | 8               |
| <i>US</i> | .20      | China        | -1.37    | 1.57         | 1.35          | 949 | .1761       | 9               |
| <i>US</i> | .32      | China        | -2.53    | 2.85         | 1.79          | 609 | .0736       | 10              |

*Note:* DGF is Differential Group Functioning measures interactions between groups of persons and groups of items (Linacre, 2018). The DGF Contrast is the difference in difficulty of the item between the two groups. This should be at least 0.5 logits for DGF to be

noticeable. Probability indicates the probability of observing this amount of contrast by chance, when there is no systematic item bias effect. DGF is powerful when looking for latent classes among the persons. In the table above, the Nation in *Italics* found the Standard harder. \* p<.0000

These data indicate that Standards 2, 5, 6, 8, 9, and 10 were more difficult for Americans (positive values), while Standards 1, 3, 4, and 7 were more difficult (negative values) for the Chinese. Of these, only Standards 3, 4, and 7 (Chinese) and 6 (US) were near or above the threshold set of one-half standard deviation (3.82 logits). Standards 4 (Content Knowledge) and 7 (Planning for Instruction) showing the most difficulty for Chinese compared to US candidates. Standard 6 (Assessment) indicated more difficulty for US candidates. The global language of these three standards follows:

- Standard 4 (more difficult for Chinese): The teacher understands the central concepts, tools of inquiry, and structures of the discipline(s) he or she teaches and creates learning experiences that make these aspects of the discipline accessible and meaningful for learners to assure mastery of the content.
- Standard 6 (more difficult for the US): The teacher understands and uses multiple methods of assessment to engage learners in their own growth, to monitor learner progress, and to guide the teacher's and learner's decision making.
- Standard 7 (more difficult for Chinese): The teacher plans instruction that supports every student in meeting rigorous learning goals by drawing upon knowledge of content areas, curriculum, cross-disciplinary skills, and pedagogy, as well as knowledge of learners and the community context.

Regarding Research Question 2, these differences are explainable. Standard 4, the mix of students (all early childhood for the Chinese but elementary and secondary for Americans) may explain the difference in commitment to content, which is less important in early childhood (the Chinese cohort). The focus on access and diversity, too, (the commitment that all students can learn) may have been more of a challenge for the Chinese students than the Americans, since access to education has traditionally been more competitive in China (Liu & Wu, 2006). Finally, the Chinese have less of a tendency to use different approaches to learning. They have maintained the traditional focus on rote learning and memorization, having found drill and practice useful in passing exams (Chan, 2004) and needing to focus attention on training students on testing practices and strategies in order to increase their test scores (Brady, 2008). Standard 6 is predictable given the high-stakes testing that is well documented in American schools (Pope & Miller, 2002). Regarding Standard 7, lesson planning has become a compulsion in American schools. The focus on structure and goals, accompanied by more focus on standards-based connections (Meier & Knoester, 2017), might have impressed this standard for Americans.

## 6.2 Findings for Individual Standards

Table 5 presents the analysis by item. Six items were extracted for explanation based on the literature, with three each showing higher commitment for one population or the other in Table 5. The six items are indicated by a double-asterisk (\*\*) under the column "BATS Item".

Table 5: Rasch Analysis by Item

| Nation | DGF SIZE | Nation | DGF SIZE | DGF CONTRAST | Rasch-Welch t | Item Degrees Freedom | Probability | BATS Item   |
|--------|----------|--------|----------|--------------|---------------|----------------------|-------------|-------------|
| US     | -2.39    | China  | 15.33    | -17.72       | -6.33         | 151                  | .0000*      | <b>01**</b> |
| US     | -1.24    | China  | 8.03     | -9.27        | -3.67         | 156                  | .0003*      | <b>02</b>   |
| US     | -1.44    | China  | 5.05     | -6.49        | -1.75         | 202                  | .0811       | 03          |
| US     | .00      | China  | -.21     | .21          | .05           | 173                  | .9595       | 04          |
| US     | .63      | China  | -3.39    | 4.02         | .63           | 158                  | .5283       | 05          |
| US     | 1.24     | China  | -9.68    | 10.91        | 4.52          | 151                  | .0000*      | <b>06</b>   |
| US     | 1.37     | China  | -10.02   | 11.40        | 1.09          | 134                  | .2780       | 07          |
| US     | -.33     | China  | 2.10     | -2.43        | -1.00         | 158                  | .3189       | 08          |
| US     | -.67     | China  | 2.54     | -3.21        | -.61          | 191                  | .5456       | 09          |
| US     | -2.42    | China  | 8.98     | -11.40       | -4.21         | 199                  | .0000*      | <b>10</b>   |
| US     | 1.31     | China  | -9.70    | 11.02        | 4.14          | 146                  | .0001*      | <b>11</b>   |
| US     | -.96     | China  | 11.39    | -12.35       | -3.50         | 126                  | .0006*      | <b>12</b>   |
| US     | 3.95     | China  | -23.80   | 27.75        | 10.96         | 174                  | .0000*      | <b>13**</b> |
| US     | .00      | China  | .00      | .00          | .00           | 175                  | 1.000       | 14          |
| US     | -.57     | China  | 8.19     | -8.76        | -2.14         | 123                  | .0344       | 15          |
| US     | .40      | China  | -3.64    | 4.04         | 1.55          | 148                  | .1228       | 16          |
| US     | 1.08     | China  | -6.68    | 7.76         | 1.24          | 146                  | .2158       | 17          |
| US     | 1.42     | China  | -10.79   | 12.21        | 4.70          | 146                  | .0000*      | <b>18</b>   |
| US     | -.27     | China  | 1.28     | -1.55        | -.54          | 172                  | .5870       | 19          |
| US     | .00      | China  | .51      | -.51         | -.16          | 173                  | .8768       | 20          |
| US     | -1.72    | China  | 5.69     | -7.41        | -1.91         | 209                  | .0580       | 21          |
| US     | .00      | China  | -.50     | .50          | .10           | 172                  | .9219       | 22          |
| US     | -2.02    | China  | 18.26    | -20.27       | -5.95         | 131                  | .0000*      | <b>23**</b> |
| US     | .00      | China  | -.71     | .71          | .22           | 168                  | .8293       | 24          |
| US     | .89      | China  | -6.02    | 6.91         | 2.61          | 151                  | .0100*      | <b>25</b>   |
| US     | 1.53     | China  | -12.01   | 13.54        | 1.30          | 129                  | .1952       | 26          |
| US     | -.70     | China  | 3.52     | -4.22        | -1.64         | 172                  | .1029       | 27          |
| US     | -.82     | China  | 3.45     | -4.27        | -1.39         | 184                  | .1660       | 28          |
| US     | 1.26     | China  | -8.35    | 9.60         | 1.55          | 141                  | .1235       | 29          |
| US     | .63      | China  | -3.39    | 4.02         | .63           | 158                  | .5283       | 30          |
| US     | -1.40    | China  | 9.51     | -10.91       | -4.20         | 152                  | .0000*      | <b>31</b>   |
| US     | -1.67    | China  | 13.01    | -14.68       | -5.12         | 142                  | .0000*      | <b>32</b>   |
| US     | -.93     | China  | 4.05     | -4.98        | -1.76         | 182                  | .0794       | 33          |
| US     | -2.78    | China  | 8.55     | -11.33       | -3.50         | 221                  | .0006*      | <b>34</b>   |
| US     | -2.63    | China  | 10.21    | -12.84       | -4.92         | 194                  | .0000*      | <b>35</b>   |
| US     | -.48     | China  | 2.02     | -2.51        | -.73          | 181                  | .4682       | 36          |
| US     | 1.71     | China  | -14.15   | 15.86        | 2.61          | 127                  | .0102*      | <b>37</b>   |
| US     | -3.05    | China  | 11.01    | -14.06       | -5.29         | 202                  | .0000*      | <b>38</b>   |
| US     | .81      | China  | -4.62    | 5.43         | 1.27          | 154                  | .2059       | 39          |
| US     | .00      | China  | 1.02     | -1.02        | -.39          | 166                  | .6961       | 40          |
| US     | .45      | China  | -2.43    | 2.88         | .81           | 162                  | .4195       | 41          |
| US     | .64      | China  | -3.66    | 4.30         | 1.30          | 157                  | .1952       | 42          |
| US     | 1.60     | China  | -12.21   | 13.81        | 4.02          | 137                  | .0001*      | <b>43</b>   |
| US     | -1.22    | China  | 6.28     | -7.50        | -3.02         | 171                  | .0029       | <b>44</b>   |
| US     | .25      | China  | -1.24    | 1.49         | .29           | 168                  | .7696       | 45          |
| US     | 3.40     | China  | -34.30   | 37.70        | 8.64          | 117                  | .0000*      | <b>46**</b> |
| US     | -3.80    | China  | 23.08    | -26.88       | -8.27         | 146                  | .0000*      | <b>47**</b> |
| US     | 1.51     | China  | -11.26   | 12.77        | 4.08          | 141                  | .0001*      | <b>48</b>   |
| US     | 3.35     | China  | -25.89   | 29.24        | 10.21         | 143                  | .0000*      | <b>49**</b> |
| US     | .68      | China  | -4.25    | 4.93         | 1.78          | 155                  | .0775       | 50          |

Note: DGF is Differential Group Functioning measures interactions between groups of persons and groups of items (Linacre, 2018). The DGF Contrast is the difference in difficulty of the item between the two groups. This should be at least 0.5 logits for DGF to be noticeable. Probability indicates the probability of observing this amount of contrast by chance, when there is no systematic item bias effect. DGF is powerful when looking for latent classes among the persons. \*p<.01

In response to Research Question 2, possible explanations for the differences follow for a representative sample of items. Items selected that were more difficult for the Chinese are 1, 23, and 47.

- 01: *We should drill on the 3Rs (reading, 'riting, and 'rithmetic) because that is the most important thing for children to learn.* (Disagree; DGF contrast of -17.32)

Guo (2005) discussed the 'Chinese Virtuoso Model' of teaching in which a teacher resembles a musician performing for the whole class. The goal is to produce an outstanding and virtuoso performance, achieved through extensive drilling and practice without interaction, questions, or expectations about individual difference. Guo recommended a shift away from traditional memorization and lecture to more student-centered and constructivist approaches, which could explain why drilling was seen as so important. Mullen (2018) suggested that for teachers working in an environment where testing is paramount, challenging the imagination and providing conditions for inventiveness is difficult. Subject matter dominates and accounts for around 7% of the curriculum for teacher training (Hu & Verdugo, 2015), so this result is understandable.

- 23: *Creativity is best taught in art and music.* (Disagree; DGF contrast of -20.27)

Mullen (2018) concluded that Chinese students' reduced focus on creativity is likely a reflection of the system and not innate capacity. She found that significant interferences with creative learning in childhood socialize adults to think they are creatively destitute, noting that this self-belief is a problem for both preservice and in-service teachers. This could explain why Chinese teacher candidates do not see their role as teaching creativity outside of the arts. The Chinese testing system is focused more on mathematics, physics, and chemistry, neglecting art and music.

- 47: *Some teachers go overboard by attending workshop, courses, meetings, and conferences to constantly learn new knowledge.* (Disagree; DGF contrast of -26.88)

Professional development is a routine matter in US schools, but, as Tian Ping (2011) noted, continued training is inefficient and viewed as not pertinent in Chinese schools. Han (2012) discussed the general lack of college level training held by teacher candidates, noting that only 23.7% of elementary teachers have an earned bachelor's degree. The lack of commitment to professional development after hire may be the result of a lack of earlier high-level coursework. Tian Ping (2011) recommended a greater focus on professional development, separation and respect for pedagogical skills vs. academic knowledge, and increased cultivation of practical skills through both pre-service, in-service, and post-graduation training. The lack of exposure to college-level coursework, combined with level of quality and experience in the trainings delivered, may explain this finding.

Items more difficult for US students include items 13, 46, and 49.

- 13: *Sometimes students may not understand why a lesson is meaningful, but that is okay as long as they learn the content.* (Disagree; DFG contrast of 27.75)  
 Han (2012) suggested that the national curriculum standards taking root in China, have reduced focus on teacher-centered teaching in the classroom. The movement provides for students learning and developing through independent thinking, inquiry and cooperation, with a transition away from memorization of 'bookish' knowledge for purely ranking or selection purposes toward a more all-round development of good character and personal attributes. The new focus on pedagogical knowledge supports this transition (Yang, 2011). The reform movement, therefore, may help to explain why Chinese teachers are more committed to helping students understand the importance of content rather than just memorizing the content. In contrast, the USA teachers have long held to constructivist approaches but have been required in recent years to focus more on answering correctly for state tests, which may explain why content has become more important than meaning for them.
- 46: *In today's world teachers are multilingual. Plan to learn another language so that I can deliver content more effectively.* (Agree; DFG contrast of 37.7)  
 Hou, Loerts and Verspoor (2018) studied the attitude and test-driven motivation towards English at Chinese universities and the belief that English is required to get better jobs. English as a foreign language (EFO) is studied from kindergarten to university with different tests administered at different levels of study, often seen as a means to an end. Thus, it is not surprising that Chinese teachers would be more committed to second language learning than Americans. The second language is the basic tool to learn from other countries and to promotion.
- 49: *In today's classroom every single lesson should have an assessment. Every assessment should be recorded and analysed.* (Agree; DFG contrast of 29.24)  
 Zhu and Han (2006) discussed the high emphasis on student test scores, with assessment a tradition dating back to Confucius. The education system in PRC places great emphasis on regular high-stakes public examinations (Brown and Wang, 2016). Although the government of China, since at least the 1990s, has introduced assessment reforms that attempt to move evaluation away from purely ranking or selection purposes towards more formative, authentic, and humanistic approaches, the degree of change in the assessment system is limited. Demand for space in higher education institutions exceeds supply, with only half of examined candidates awarded a place, standing in sharp contrast to the USA system of open access. The centuries old culture of assessment in China can explain this result.

## 7. Conclusions and Limitations

The differential group functioning (DGF) Rasch procedure revealed specific differences between Chinese and American teacher education students at the

InTASC standards level and at the individual items level. Three standards (4, 6, and 7) and six items (3 per group) were targeted for analysis based on their DGF contrast. A judgmental analysis was provided for each, treating the items as 'cases.' The differences were consistent with the literature.

With the Chinese standards for teaching being similar to the US InTASC Standards (with the exception of the emphasis on content knowledge and instruction), the training should be similar enough to make these two groups comparable. The results, however, indicated otherwise, lending support to the conclusion that culture and nationality are possible explanations for the differences in levels of commitment to the critical dispositions of teaching.

At a more specific level, Chinese teacher preparation programs could explore whether the weaker commitment to content and planning is a result of the area of concentration of the students sampled or a more generalizable issue. US programs should focus on improving the commitment of teacher candidates to assessment, helping them to understand the importance of formative assessment in particular. The decreased focus on creativity and professional development among Chinese could be an area of focus for their teachers. Similarly, US teacher preparation programs could focus on the need for students to understand the importance of what they are learning, the importance of being multilingual, and the need for continuous assessment.

Regarding limitations, convenience sampling limits generalizability. The Chinese cohort responded to the survey written in English, so there may be language loading. There has been no crosswalk between InTASC and the Chinese NCSTE, so there may be differences in what was taught by design.

## **8. Implications and Future Research**

Rasch standards for reliability continue to be met, and construct validity is supported by the logical and literature-based analysis of the results, particularly at the item level, as hoped. This provides promise for continued and expanded use of the instrument. The study provides evidence that American teacher educators are teaching what is expected, which can be celebrated.

There are possible explanations yet to be explored in future research. For example, exploration in dispositional behaviors that support student learning and development are important ways of interacting that undergird program development and professional interactions with students, colleagues, parents, and communities (Moore et al., 2019, a & b). The interactive differences in culture may be the best explanation for the comparative differences of admitted teacher candidates, but it may also be that training in the USA is improving teaching dispositions (LaPaglia, 2020).

Recommendations for future research include:

1. Examine the training on dispositions received in both countries and how it is related to the InTASC standards coincidentally or intentionally.
2. Identify a population of Chinese teacher candidates planning to teach at the elementary and secondary levels to add to the dataset.

3. Use additional measures from the DAATS battery to explore more deeply the commitments of both populations.
4. Translate BATS2 into Mandarin and replicate the study.

At a policy level, particularly in the US, standards authors might look at how the standards compare and, most specifically why one culture places dispositions first (China) and the other places it last (US), remembering where this article began – the memory most US adults have of their best teacher is dispositions-based.

## 9. References

- Almerico, G., Johnston, P., Henriott, D., & Shapiro, M. (2011). Dispositions assessment in teacher education: Developing an assessment instrument for the college classroom and the field. *Research in Higher Education Journal*, 11, 1-19. Retrieved from <https://www.aabri.com/manuscripts/10665.pdf>
- Bair, M. A. (2017). Identifying dispositions that matter: Reaching for consensus using a Delphi Study. *Teacher Educator*, 52(3), 222-234. <https://doi.org/10.1080/08878730.2017.1315475>
- Behrmann, L., & Souvignier, E. (2013). Pedagogical content beliefs about reading instruction and their relation to gains in student achievement. *European Journal of Psychology of Education*, 28(3), 1023-1044. <https://doi.org/10.1007/s10212-012-0152-3>
- Boonen, T., Van Damme, J., & Onghena, P. (2014). Teacher effects on student achievement in first grade: Which aspects matter most? *School Effectiveness and School Improvement*, 25(1), 126-152. <https://doi.org/10.1080/09243453.2013.778297>
- Brady, A.L. (1999). *Effects of standardized testing on teachers' emotions, pedagogy and professional in interactions with others* [Doctoral dissertation, Cleveland State University]. Retrieved from <https://engagedscholarship.csuohio.edu/etdarchive/39>.
- Chan, K. (2004). Preservice teachers' epistemological beliefs and conceptions about teaching and learning: Cultural implications for research in teacher education. *Australian Journal of Teacher Education*, 29(1), 1-3. <https://doi.org/10.14221/ajte.2004v29n1.1>
- Council for the Accreditation of Educator Preparation. (2016). *The CAEP Standards*. Retrieved from <http://caepnet.org/~media/Files/caep/standards/caep-standards-one-pager-0219.pdf?la=en>
- Council of Chief State School Officers. (2013). InTASC Standards. Retrieved from <https://ccsso.org/resource-library/intasc-model-core-teaching-standards>
- Cox, J. L., Cheser, K., & Detwiler, J. (2015). *A triptych study of the impact of teacher dispositions on teacher hiring and student outcomes, teacher and student growth mindsets, and student perceptions of teacher-student relationships* (UMI No. 3707252). [Doctoral dissertation, Northern Kentucky University]. ProQuest Dissertations and Theses Global.
- Eberly, J. L., Rand, M. K., & O'Connor, T. (2007). Analyzing teachers' dispositions towards diversity: Using adult development theory. *Multicultural Education*, 14(4), 31-36.
- Eckes, T. (2015). *Introduction to many-FACET Rasch measurement: analysing and evaluating rater-mediated assessment (2nd)*. New York: Peter Lang. <https://doi.org/10.3726/978-3-653-04844-5>
- Edgington, W. D., & Cox, M. (2015). Implementing professional dispositions and behavior with preservice teachers: One program's journey. *National Teacher Education Journal*, 8(3), 47-54.
- Englehart, D. S., Batchelder, H. L., Jennings, K. L., Wilkerson, J. R., Lang, W. S., Quinn, D. (January 2012). Teacher dispositions: Moving from assessment to improvement.

- The International Journal of Educational and Psychological Assessment*, 9(2), 26-44. Retrieved from [http://tijepa.books.officelive.com/Documents/V9.2\\_TIJEPA.pdf](http://tijepa.books.officelive.com/Documents/V9.2_TIJEPA.pdf)
- Gavin, T. L., & Wang, Z. (2016). Understanding Chinese university student conceptions of assessment: Cultural similarities and jurisdictional differences between Hong Kong and China. *Social Psychological Education*, 19(5), 151-173. <https://doi.org/10.1007/s11218-015-9322-x>
- Han, (2012). Big moves to improve the quality of teacher education in China. *On the Horizon*, (20)4, 324-335. <https://doi.org/10.1108/10748121211272461>
- Hou, J., Loerts, H., & Verspoor, M. (2018). Exploring attitude and test-driven motivation towards English at Chinese universities. *International Journal of Language Studies*, 12(1), 37-60.
- Hu, A., & Verdugo, R. (2015). An analysis of teacher education policies in China. *International Journal of Educational Reform*, (24)1, 37-54. <https://doi.org/10.1177/105678791502400104>
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1973). *Taxonomy of educational objectives, the Classification of educational goals. Handbook II: Affective domain*. New York: David McKay Co., Inc. <https://doi.org/10.1177/002248716501600228>
- Lang, W. S., & Wilkerson, J. R. (2006). *Measuring teacher dispositions systematically using INTASC Principles: Building progressive measures of dispositions*. [Paper presentation]. American Association of Colleges of Teacher Education, San Diego, CA.
- Lang, W. S., & Wilkerson, J. (March, 2008). *Measuring teacher dispositions with different item structures: An application of the Rasch Model to a complex accreditation requirement*. [Paper presentation]. American Education Research Association, New York.
- Lang, W. S., & Wilkerson, J. (2008). Measuring teacher dispositions with different item structures: An application of the Rasch Model to a complex accreditation requirement. *Resources in Education*. Retrieved from <https://files.eric.ed.gov/fulltext/ED502965.pdf>
- Lang, W. S., Moore, L., Wilkerson, J. R., Parfitt, C. M., Greene, J., Kratt, D., Martelli, C. D., LaPaglia, K., Johnston, V., Gilbert, S., Zhang, J. (2018). Beliefs About Teaching (BATS2) – Construction and validation of an instrument based on InTASC critical dispositions. *International Journal of Learning, Teaching and Educational Research*, (17)8, 56-77. <https://doi.org/10.26803/ijlter.17.8.4>
- Lang, W. S., Wilkerson, J. R., Moore, L. L., Fields, L. J., Parfitt, C. M., Greene, J. S., Kratt, D. M., Martelli, C. D., LaPaglia, K. E., Johnston, V. D., Gilbert, S. G., Zhang, J., & Wang, C. X. (2018). *Measuring teacher dispositions systematically using multiple measures*. [Paper presentation]. Eastern Educational Research Association, Clearwater, FL.
- Lang, W. S., Wilkerson, J. R., Rea, D. C., Quinn, D., Batchelder, H. L., Englehart, D. S., & Jennings, K. J. (2014). Measuring teacher dispositions using the DAATS battery: A multifaceted Rasch analysis of rater effect. *Journal of Applied Measurement*, 15(3), 240-251.
- LaPaglia, K. (2020). *Preservice teacher dispositions: A case study of changes during a teacher preparation program*. [Unpublished doctoral dissertation]. Florida Gulf Coast University.
- Linacre, J. M. (2018). *Winsteps* (Version 3.93.2) [Computer software]. Chicago, IL. Retrieved from <https://www.winsteps.com/index.htm>
- Liu, H., & Wu, Q. (2006). Consequences of college entrance exams in china and the reform challenges. *KEDI Journal of Educational Policy*, 3(1), 7-21. Retrieved from <https://www.kedi.re.kr/eng/kedi/bbs/B0000005/list.do?menuNo=200067>



- Meier, D., & Knoester, M. (2017). *Beyond testing: Seven assessments of students and school more effective than standardized tests*. New York: Teachers College Press.
- Moore, L. L., Lang, W. S., Wilkerson, J. R., Parfitt, C., & LaPaglia, K. (October, 2019a). *I love to teach but, based on my personal dispositions, will I be able to connect to my students and persist in the profession?* [Roundtable presentation] Southern Regional Educational Leaders Association, Ft. Myers, Fl.
- Moore, L. L., Lang, W. S., Wilkerson, J. R., Parfitt, C., & LaPaglia, K. (October, 2019b). *Leadership challenges: Changing dispositions of urban teachers from unaware to valuing*. [Roundtable presentation] Southern Regional Educational Leaders Association, Ft. Myers, Fl.
- Mullen, C. A. (2018). Creative learning: Paradox or possibility in China's restrictive preservice teacher classrooms? *Action in Teacher Education*, (4)2, 186-202. <https://doi.org/10.1080/01626620.2018.1424054>
- Pope, C., & Miller, S. (2002). Conversations from the commissions: reflective teaching in the panic of high-stakes testing. *English Education*, (34)2, 164-168. <https://library.ncte.org/journals/ee/issues/v34-2>
- Popham, W. J. (2020). *Classroom Assessment: What Teachers Need to Know, 9<sup>th</sup> ed.* Boston, MA: Pearson, Inc.
- Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. (Copenhagen, Danish Institute for Educational Research), expanded edition (1980) [foreword & afterword Wright, B. D.]. Chicago, IL: The University of Chicago Press.
- Schulte, L., Edick, N., Edwards, S., & Mackiel, D. (2004). The development and validation of the Teacher Dispositions Index. *Essays in Education*, 12, 1-16. Retrieved from [https://ehhp.cofc.edu/ncate\\_2012/documents/schulte.pdf](https://ehhp.cofc.edu/ncate_2012/documents/schulte.pdf)
- Singh, D. K., & Stoloff, D. L. (2008). Assessment of teacher dispositions. *College Student Journal*, 42(4), 1169.
- Smith, R. M. & Wind, S. A. (2018). *Rasch measurement models: interpreting WINSTEPS and FACETS output (2<sup>nd</sup>)*. Maple Grove, MN: JAM Press.
- Thurstone, L. L. (1928). Attitudes Can Be Measured. *American Journal of Sociology*, 33(4), 529-554. <https://doi.org/10.1086/214483>
- Tian-Ping, Y. (2011). The problems and some suggestions in the current system reform of teacher education in China. *US-China Education Review B1*, 117-125.
- Tian-Ping, Y. (2012). The retrospection and elicitation of China's teacher education reform and opening-up more than 30 years. *US-China Education Review B1*, 88-98. <https://doi.org/10.17265/2161-6248/2012.01B.010>
- Vanes, M. (2012) Identifying Country-Specific Cultures of Physics Education: A differential item functioning approach. *International Journal of Science Education*, 34(16), 2483-2500. <https://doi.org/10.1080/09500693.2012.684804>.
- Vaughn, K. A. (2012). *Teacher dispositions and student achievement* (UMI No. 3505422) [Doctoral dissertation]. ProQuest Dissertations and Theses Global database.
- Wasicsko, M. M. (2007). The perceptual approach to teacher dispositions: The effective teacher as an effective person. In M. E. Diez & J. Raths (Eds.), *Dispositions in teacher education* (pp. 55-91). Charlotte, NC: Information Age Publishing.
- Wilkerson, J. R. (2006, August). Measuring teacher dispositions: Standards-based or morality-based? *Teachers College Record*. Retrieved from <https://www.tcrecord.org/Content.asp?ContentId=12493>
- Wilkerson, J. R. (2012). Measurement and evaluation perspectives on scaling teacher affect with multiple measures. *The International Journal of Educational and Psychological Assessment*, 9(2), 165-191.

- Wilkerson, J. R. & Lang, W. S. (2006). Measuring teacher ability with the Rasch Model by scaling a series of product and performance tasks. *Journal of Applied Measurement*, 7(3), 239-259.
- Wilkerson, J. R., & Lang, W. S. (2007). *Assessing teacher dispositions: Five standards-based steps to valid measurement using the DAATS model*. Thousand Oaks, CA: Corwin Press.
- Wilkerson, J. R., & Lang, W. S. (2011). Standards-based teacher dispositions as a necessary and measurable construct. *The International Journal of Educational and Psychological Assessment*, 7(2), 34-54.
- Yu (2013). *Teacher Education in China: Current situation and related Issues*. Zhejiang Normal University. Retrieved from <http://www.uta.fi/cerec/educationandresearch/ChinaEduLecture/Teacher%20Education%20in%20China.pdf>
- Zhu, X. and Han, H. (2006). Reconstruction of the teacher education system in China. *International Education Journal* 7(1), 66-73. <https://eric.ed.gov/?id=EJ847303>