Published in final edited form as:

Read Writ. 2013 July 1; 26(6): 1031–1056. doi:10.1007/s11145-012-9405-4.

Developmental and Individual Differences in Chinese Writing

Connie Qun Guan, Feifei Ye, Richard K. Wagner, and Wanjin Meng

University of Science and Technology Beijing, University of Pittsburgh, Florida State University, Florida Center for Reading Research

Abstract

The goal of the present study was to examine the generalizability of a model of the underlying dimensions of written composition across writing systems (Chinese Mandarin vs. English) and level of writing skill. A five-factor model of writing originally developed from analyses of 1st and 4th grade English writing samples was applied to Chinese writing samples obtained from 4th and 7th grade students. Confirmatory factor analysis was used to compare the fits of alternative models of written composition. The results suggest that the five-factor model of written composition generalizes to Chinese writing samples and applies to both less skilled (Grade 4) and more skilled (Grade 7) writing, with differences in factor means between grades that vary in magnitude across factors.

Keywords

Chinese writing; Individual differences; Developmental differences; Chinese

Writing is a complex process that develops over a long time period. A partial list of activities that can be involved in writing includes pretask planning, online planning, idea generation, translation, transcription, text generation, revision, meeting goals for content and grammaticality, as well as retrieving words and organizing these words into meaningful language and text (McCutchen, 1996). An early model of writing proposed by Hayes and Flower (1980) and updated by Hayes (1996) organized writing activities such as these into the categories of planning, translation, and review. Berninger and Swanson (1994) subsequently proposed dividing translation into text generation, which refers roughly to putting one's ideas into words, and transcription, which refers to getting the words on paper.

Although still in its infancy compared to research on reading, a substantial literature has developed on aspects of writing. Areas of research activity include writing measurement, normal development, underlying processes, writing problems, and teaching and intervention (see, e.g., Berninger, 2009; Fayol, Alamargot, & Berninger, in press; Graham & Harris, 2009; Greg & Steinberg, 1982; Grigorenko, Mambrino, & Priess, 2011; Levy & Ransdell, 1996; MacArthur, Graham, & Fitzgerald, 2006).

When individuals are asked to write, inspection of what they produce reveals two obvious facts about writing. First, developmental differences are pronounced (McCutchen, 1996). Older advanced writers produce much longer and more complex writing samples than do younger beginning writers. Second, within a developmental level, individual differences in writing are pronounced. Some individuals are much better writers than others. One approach that has proven to be successful in analyzing developmental and individual differences in various cognitive domains has been to attempt to identify underlying factors or dimensions that account for these differences (Hooper et al. 2011).

An example of applying this approach to the domain of writing is provided by Puranik, Lombardino, and Altmann (2008), who analyzed writing using a retelling paradigm in which students listened to a story and then wrote what they remembered. The writing samples were transcribed into a database using the *Systematic Analysis of Language Transcript* (SALT) (Miller & Chapman, 2001) conventions. Although developed originally for analysis of oral language samples, its adaptation to analysis of writing samples has provided a systematic approach for coding variables (Nelson, Bahr & Van Meter, 2004; Nelson & Van Meter, 2002, 2007; Scott & Windsor, 2000). Puranik et al. (2008) used exploratory factor analysis to analyze their writing samples and interpreted a three-factor solution as representing productivity, complexity, and accuracy. Because SALT was developed for analysis of oral language samples rather than for writing using a specific orthography, a potential advantage of SALT coding for analyzing written language samples across different orthographies, is that its codes reflect aspects of language that are likely to be general across languages as opposed to writing-system specific conventions.

More recently, Wagner et al. (2011) used confirmatory factor analysis to compare models of the underlying factor structure of writing samples provided by first- and fourth-grade students. This study replicated and extended the Puranik et al. (2008) study by analyzing writing to a prompt as opposed to story retelling, using confirmatory factor analysis to test apriori specified models, representing higher-level or macro-structural aspects of text, and including measures of handwriting fluency. Handwriting fluency was included because it has been shown to be an important predictor of composition in previous studies (Graham, Berninger, Abbott, Abbott, & Whitaker, 1997). The writing samples were coded using SALT conventions.

An identical five-factor model provided the best fit to both the first- and fourth-grade writing samples. The factors were complexity, productivity, spelling and pronunciation, macro-organization, and handwriting fluency. Handwriting fluency was related not only to productivity but also to macro-organization for both grades. Correlations between handwriting fluency and both the quality and length of writing samples have been noted previously (Graham et al., 1997). The reason that handwriting fluency is related to written composition has yet to be determined definitively. One explanation that has received some empirical support is that being fluent in handwriting frees up attention and memory resources that can be devoted to other aspects of composition (Alves, Castro, Sousa, & Stromqvist, 2007; Chanquoy & Alamargot, 2002; Christensen, 2005; Connelly, Campbell, MacLean, & Barnes, 2006; Connelly, Dockrell, & Barnett, 2005; Dockrell, Lindsay, &

Connelly, 2009; Graham et al., 1997; Kellogg, 2001, 2004; McCutchen, 2006; Olive, Alves, & Castro, in press; Olive & Kellogg, 2002; Peverly, 2006; Torrance & Galbraith, 2006).

Skilled writing requires automaticity of low-level transcription and high-level construction of meaning for purposeful communication (Berninger, 1999). According to the simple view of writing (Berninger, 2000; Berninger & Graham, 1998), developing writing can be represented by a triangle in a working memory environment in which transcription skills and self-regulation executive functions are at the base that enable the goal of text generation at the top (Berninger & Amtmann, 2003).

Automaticity is achieved when a given process can be carried out accurately, swiftly, and without a need for conscious attention (LaBerge & Samuels, 1974). Berninger and Graham (1998) stress that writing is “language by hand” and point out that their research suggests that orthographic and memory processes (i.e., the ability to recall letter shapes) contribute more to handwriting than do motor skills (Berninger & Amtmann, 2003). That is to say, handwriting is critical to the generation of creative and well-structured written text and has an impact not only on fluency but also on the quality of writing (Berninger & Swanson, 1994; Graham et al., 1997). Lack of automaticity in orthographic-motor integration can seriously affect young children’s ability to express ideas in text (Berninger & Swanson, 1994; Connelly & Hurst, 2001; De La Paz & Graham, 1995; Graham, 1990; Graham et al., 1997).

Two important alternative views of the factor structure of written composition should be mentioned. The first is a levels of language framework in which the key distinctions are between the word, sentence, and text levels (Abbott, Berninger, & Fayol, 2010; Whitaker, Berninger, Johnston, & Swanson, 1994). Within this framework, the Wagner et al. (2011) productivity factor could be considered a word-level factor, the complexity factor can be considered a sentence-level factor, and the macro-organization factor can be considered a text-level construct. The second alternative view is that writing and reading both represent the same unidimensional construct (Mehta, Foorman, Branum-Martin, & Taylor, 2005). Mehta et al. scored writing samples by rating them on eight dimensions that were then combined into a single writing ability estimate. When the data were modeled at both the level of the student and the level of the classroom, the writing ability estimate and a reading ability estimate loaded on the same factor.

Chinese writing systems and writing research

Much of the existing research has been limited to the study of writing in English. To contribute to expanding knowledge of writing beyond English, the present study focused on written compositions provided by students in China.

English is an alphabetic writing system in which phonemes correspond to functional spelling units (usually one or two letters); the same phoneme can correspond to a small set of alternative one-or two-letter functional spelling units referred to an alternation (Venesky, 1970; 1999). Thus, spelling in English is a phonological-to-orthographic translation. In contrast, Chinese script is non-alphabetic and a Chinese graph is basically morphosyllabic (Lui, Leung, Law, & Fung, 2010), in which most symbols represent words or morphemes

rather than having a grapheme-phoneme correspondence. Compared with English, the pronunciation of the Chinese characters is not transparent, and grapheme (or basic graphic units corresponding to the smallest segments of speech in writing) simultaneously encode the sounds and meaning at the syllable level (Coulmas, 1991; DeFrancis, 2002; Shu & Anderson, 1999).

Furthermore, the characters or symbols of Chinese writing may represent quite different-sounding words in the various dialects of Chinese, but they represent specific form and meaning. The character is the building block for multi-morphemic words, and characters can be combined to form multipart or compound words and derivatives (Hoosain, 1991; Ju & Jackson, 1995).

When learning to write, Chinese children usually start from stroke writing, then progress to radical (the combination of several strokes) writing, and finally to whole character writing. The relation between meaning and its representation in writing is emphasized not only on a radical level and a whole character level, but also on a two character compound word level. Therefore, repeated practice with writing is commonly used to strengthen associations among orthography, semantics, and finally phonological aspects of Chinese (Guan, Liu, Chan, & Perfetti, 2011). The theoretical rationale for this type of writing practice is based on differences between languages. In contrast to the alphabetic languages, access to an orthographic entry in Chinese does not necessitate prior access to a phonological word form, but can be accessed from a semantic representation directly without phonological mediation (e.g., Rapp, Benzing, & Caramazza, 1997). In other words, although it is correct to assume rules to convert phonemes to grapheme in alphabetic languages (e.g., Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), graphemes do not exist in Chinese and so there is no reason to assume any equivalent correspondences between sound and spelling (Weekes, Yin, Su, & Chen, 2006). This implies that language specific mapping between other types of representations in Chinese might be used for writing (stroke, radicals, rime, tones). Indeed, literacy in Chinese emphasizes the role of strokes, radicals and whole characters in handwriting (Perfetti & Guan, 2012).

Most writing research in Chinese has focused on Chinese character acquisition (Guan et al., 2011; Lin et al., 2010) and character recognition (Ju & Jackson, 1995; Leck, Weekes, & Chen, 1995; Perfetti & Zhang, 1995; Shu & Anderson, 1999; Weekes, Chen, & Lin, 1998). Unlike issues for the English language that have been widely studied, less is known about written composition in Chinese.

One exception is a recent study by Yan et al. (in press). They examined written composition among elementary school students in Hong Kong. They developed an index of overall writing quality that was based on summing together five variables, each of which was rated on a 1- to 4-point scale. Depth was a rating of the extent to which the ideas were elaborated. Sentence-level organization was a rating of whether sentences were complete and connectives and sequencers were used. Paragraph-level organization was a rating of the extent to which the organizational structure of the passage was effective for conveying the intended meaning. Prominence of organizational or key elements was a rating of the extent to which topic sentences and concluding sentences were used appropriately. Finally,

intelligibility was a rating of the extent to which the writing sample was easy to understand and pleasant to read.

There were two key results from this study. First, a single underlying factor explained individual differences on the five variables that were rated, which supported combining them into a single overall score. Thus, writing performance was captured by a single factor rather than multiple factors. Second, predictors of the measure of overall writing quality included vocabulary knowledge, Chinese word dictation skill, phonological awareness, speed of processing, speeded naming, and handwriting fluency.

The present study

The goal of the present study was to examine the generalizability of the five-factor model (Wagner et al., 2011) of the underlying dimensions of written composition across writing systems (Chinese Mandarin vs. English) and level of writing skill. There were two specific reasons for using the five-factor model as opposed to other possible models in the present study. First, the five-factor model addresses developmental and individual differences in writing, which were of interest in the present study. Second, because the model was implemented as a confirmatory factor analytic model, it was possible to conduct a relatively rigorous test of the fit of the model to Chinese writing samples compared to other models of writing that have not been implemented as confirmatory factor analytic models.

For the present study, Chinese writing samples were obtained from 4th and 7th grade students. The rationale for choosing grade 4 and 7 participants in this study was to both match a grade level used in Wagner et al. (2011) (grade 4) and to extend the study of writing samples to a higher grade level (grade 7). In addition, Chinese students are beginning to receive a formal writing course at grade 4, and in grade 7 their writing training becomes more intensive and systematic.

Confirmatory factor analysis was used to examine the fit of the five-factor model to the data. Our major research question was to determine which aspects of the five-factor model of written composition that was developed from analyses of English writing samples would apply to Chinese writing samples. Although the results of Yan et al. (in press) suggest that quality of Chinese writing might be unidimensional, their data were quality ratings on 1- to 4-point scales, as were the English data of Mehta et al. (2005) that also supported a unidimensional model. Specifically, by modeling quantitative variables in Chinese writing samples that were comparable to those obtained by Wagner et al. (2011) as opposed to quality ratings, we attempted to determine whether a multi-factor model of writing would fit the data when writing is analyzed by quantitative variables rather than quality ratings.

Second, one surprising finding in the Wagner et al. (2011) analyses of English writing samples was that the same five-factor model fit the data from writing samples provided by first- and fourth-grade students. Therefore, our secondary research question was to examine whether the identical five-factor model would apply to writing samples provided by more advanced writers. This was addressed by analyzing the data provided by seventh-grade writers as compared to fourth-grade writers.

Finally, in the previous study, only a single writing prompt was used to obtain the writing samples that were analyzed. In the present study, the third research question was related to the stability of parameters of the model. Writing samples obtained from two writing prompts were analyzed to examine the stability of parameters of the model across writing samples produced to different writing prompts.

METHODS

Participants

Writing samples were collected from 160 Grade 4 students and 180 Grade 7 students from one typical primary school and one middle school in Beijing. For Grade 4 students, there were 85 boys (53.1 %) and 75 girls (46.9 %) with an average age of 10.1 years. For Grade 7 students, there were 92 boys (50.8 %) and 88 girls (49.2 %) with an average age of 13.3 years. Socioeconomic status of the students was primarily middle and lower class. All the students at the primary and middle schools were speaking putonghua, a standard Beijing dialect.

Measures

The measure consisted of two compositional writing samples and two handwriting fluency measures.

Writing samples—The writing samples were obtained using two counterbalanced prompts.

Prompt 1: We are going to write about selecting a student as our class monitor. Imagine you are going to elect only one student as your class monitor. Who will that student be? Why do you want to elect this student as your class monitor?

Prompt 2: We are going to write about choosing a gift for your mother. Imagine you are going to select only one gift to give to your mother. What will that gift be? Why do you want to choose that gift for your mother?

Both prompts were introduced by saying: “When you are writing today, please stay focused and keep writing the whole time. Don’t stop until I tell you to do so. Also if you get to a character that you don’t know how to spell, do your best to write it out by using a character with similar sound or a character with similar form. I’m not going to help you with character writing today. If you make a mistake, cross out the character you don’t want and keep writing. Don’t erase your mistake because it will take too long. Keep writing until I say stop. You will have a total of 10 min for completing writing on this topic”.

The rationale for selecting the specific writing prompts was to encourage students to think creatively and write something that they are capable of writing. The prompts were relevant to students’ daily life experiences, so that the students should all have something to say about the topics. Both prompts required the students to present some reasons to support their opinions.

Written samples were hand coded using Systematic Analysis of Language Transcript conventions (*SALT*, Miller & Chapman, 2001) by the first author and three graduate students. Detailed description of each of these ten *SALT* variables is given below. They were organized into four tentative constructs for the subsequent confirmatory factory analytic modeling:

Macro-organization

- 1 *Topic*. A score of 1 or 0 was given to indicate whether the written sample included a topic sentence or not.
- 2 *Logical ordering of ideas (Order)*. A 1- to 4-point rating scale was used to assess the logical ordering of idea of the students' written sample.
- 3 *Number of key elements*. One point each was given to assess whether the written sample include a main idea, a main body, and a main conclusion of the content, thus yielding to a maximum of 3 points in total.

Complexity

- 4 *Mean length of T-unit (MLT)*. The total number of characters in students' composition divided by the total number of T-units.
- 5 *Clause Density (CD)*. The total number of characters in students' composition divided by the total number of clauses.

Productivity

- 6 Total number of characters (TNC).
- 7 Total number of different characters (NDC).

Spelling and punctuation (mechanical errors)

- 8 *Number of alternative characters which have the similar pronunciation or homophone (PHE)* as the target character, e.g., “诞” in “圣诞 (Shèngdàn, target)”-“旦” in “圣诞 (Shèngdàn)”
- 9 *Number of alternative characters which have a similar orthographic form (ORE)* of the target character, e.g., “诞” in “圣诞 (Shèngdàn, target)”-“延” in “圣诞 (Shèng yán)”
- 10 *Number of errors involving punctuation (PNE)*.

The third author trained all the research assistants in *SALT* coding. The first author and three graduate students coded all writing samples when they were familiarized with the coding rubrics after practicing. Each written sample was coded twice. Disagreement was solved by discussion. We calculated inter-rater reliability based upon randomly selected written samples. Twenty-five percent of the writing samples were randomly selected, with 5 to 6 students' two-passage essays chosen from each of six classes. Inter-rater reliability was assessed for the above-mentioned ten variables. The inter-rater reliability ranged from 75 to 100 % for coded items across transcripts.

Handwriting fluency tasks—Handwriting fluency was assessed by a stroke copying fluency task and a sentence copying fluency task. Following the same rationale and implementation in Wagner et al. (2011), these tasks required the students to demonstrate their ability to write single strokes or single characters as well and as quickly as they can. Both tasks were introduced to the participants to play a game of copying tasks. The first task asked them to copy varied single strokes line by line. There were five lines of strokes with ten single strokes on each line (e.g., ノ ㇏ ㇏ ㇏ ㇏ ㇏ ㇏ ㇏ ㇏ ㇏). Each line was composed of a random selection of 10 strokes out of a total of 30 varied strokes. The participants were given 60 s to copy down as many strokes as possible. We randomized the order of the strokes to avoid students memorizing the stroke order, thus the copying speed is purely determined by the students' single-stroke copying ability. The scoring of this task was the total number of strokes written within 60 s. The test–retest reliability of this stroke copying fluency task was .93.

The second task asked the participants to copy one sentence, e.g., 敏捷的棕狐狸跳越懒狗 (in English translation: A quick brown fox jumped over the lazy dog). There was a total of 10 Chinese characters in this sentence. This task followed the same rationale with the first stroke-copying task, i.e., all of the characters contained almost the full range of single strokes. In 60 s, the participants were required to copy this 10-character sentence as many times as they can. No linkage of strokes between characters was allowed so as to make each character as a stand-alone one as they wrote. The total score of this task was the sum of single characters correctly copied in order. The test–retest reliability of this sentence copying fluency task is .91.

Procedure

All the students were assessed in twelve classes by their Chinese instructors, who administered the test along with the experimenters at the same time during the normal 45 min class period. All the instructions were audio-taped and played by the loudspeaker to the students at the same time to all twelve classes. All tasks were group administered in this way.

The twelve classes followed the same time constraint and experimental schedule. In each class, there was one experimenter and one Chinese instructor monitoring task administration and to answer students' questions in related to all assessments during the study.

Half of the students were asked to complete one of the written essays first, and then to complete a second written essay later. There were 2 min breaks given between the two writing assignments. Immediately after the writing tasks, the students were given handwriting fluency tasks, with stroke copying fluency task first, and sentence copying fluency task second. Demographic information was also collected.

Data analysis plan

The data analysis was carried out in two steps after data screening. In the first step, four separate CFA models were analyzed to test the proposed five-factor factorial structure for each writing sample (A and B) and grade (4 and 7). For each CFA model, one of the factor

loadings for each factor was fixed to be one for model identification. In the second step, we assessed measurement invariance across writing samples and grades separately. The purpose of testing measurement invariance was to establish that either partial- or full-measurement invariance was established across writing sample and grade. Failing to do so would preclude meaningful comparisons across writing samples or grades because of concern that the latent variables were not comparable. For the test of measurement invariance across grades, multi-group CFA were used. For the test of measurement invariance across writing samples, multi-group CFA would not have been appropriate here because writing samples A and B were administered to the same subjects. This analysis was done in single-group CFA models that included both writing samples. A stepwise procedure was adopted to assess measurement invariance (Vandenberg & Lance, 2000): (1) A baseline model was analyzed without any equality constraints for corresponding factors; (2) an equal factor loading model was analyzed with equality constraints imposed on corresponding factor loadings. If all factors' loadings were invariant, we continued to (3) assess invariance of intercept. If all factor loadings were not invariant, we found out which variables had equal factor loadings and then among these variables, which had equal intercepts. The Chi-square difference test was used to assess the invariance of factor loadings and intercepts. Chi-square difference testing was conducted using the Satorra-Bentler adjusted Chi-square (Satorra, 2000; Satorra & Bentler, 1988).

The goodness of fit between the data and the specified models was estimated by employing the Comparative Fit Index (CFI) (Bentler, 1990), the TLI (Bentler & Bonett, 1980), the RMSEA (Browne & Cudeck, 1993), and the standardized root mean squared residual (SRMR; Bentler, 1995). CFI and TLI guidelines of greater than 0.95 were employed as standards of good fitting models (Hu & Bentler, 1999). Different criteria are available for RMSEA. Hu and Bentler (1995) used .06 as the cutoff for a good fit. Browne and Cudeck (1993) and MacCallum, Browne, and Sugawara (1996) presented guidelines of assessing model fit with RMSEA: values less than .05 indicate close fit, values ranging from .05 to .08 indicate fair fit, values from .08 to .10 indicate mediocre fit, and values greater than .10 indicate poor fit. A confidence interval of RMSEA provides information regarding the precision of RMSEA point estimates and was also employed as suggested by MacCallum et al. (1996). ASRMR <.08 indicates a good fit (Hu & Bentler, 1999). All CFA and measurement invariance analysis were performed with Mplus 6.1 (Muthén & Muthén, 2010).

RESULTS

Data screening

Table 1 presents the descriptive statistics by grade and writing sample. Because of minimal variability in whether a topic sentence was present, this variable was combined with the number of key elements. Tables 2 and 3 present bivariate correlations among the twelve variables for grades 4 and 7 respectively. These correlations suggest that these variables are moderately correlated.

We screened the raw data for normality, and due to some departure from multivariate normality, we adopted robust maximum likelihood estimation (MLR in Mplus). For non-

normal data, this estimation procedure functions better than maximum likelihood (Hu, Bentler, & Kano, 1992).

We found that the missing data patterns across groups were proportionately similar, which suggests that missing data were missing completely at random. Students with missing responses on some items were retained for analysis by using direct maximum likelihood estimation with missing data in Mplus 6.1 (Kline, 2011).

Confirmatory factor analysis

Confirmatory factor analysis was carried out separately on the two grade 4 and the two grade 7 writing samples. Table 4 presents model fit indices. The five-factor model had an adequate fit for grade 4 writing samples and an excellent fit for grade 7 writing samples. Figures 1, 2, 3, and 4 present standardized factor loadings and inter-factor correlations by grade and writing sample. Number of period errors was not significantly loaded on the factor of spelling and punctuation for both writing samples at both grades, and thus was deleted from further analysis. This makes sense because Chinese punctuation tends to be quite free-flowing and more ambiguous than English with regard to positioning of commas and periods.

Measurement invariance—We examined the measurement invariance between writing sample A and writing sample B for grade 4. We employed a CFA with the writing sample A variables loaded on the latent factors corresponding to writing sample A and the writing sample B variables loaded on the latent factors corresponding to writing sample B. Given that the same manifest variables were used for both writing sample A and writing sample B, residuals of the corresponding variables were first allowed to be correlated and then excluded from the final model when found insignificant. For the factor of handwriting fluency, the manifest variables have the same values for writing samples A and B, thus creating singularity in the covariance matrix. We did not include this factor when examining measurement invariance. The model fit of the restrictive model constraining the factor loading to be the same for the corresponding variables were compared against the unrestrictive model with no such constraints. Two measures had correlated residuals across writing sample A and B, the Topic + Number of key elements ($r = .31, p < .001$), and number of different characters ($r = .34, p < .001$).

The model fit and Chi-square difference tests are presented in Table 5. The baseline model provided a good fit $\chi^2_{(df=77)}=125.17, p < .001$, CFI = .97, TLI = .95, RMSEA = .06 (90% CI .04–.08), and SRMR = .07. The restrictive model with equal loadings had an adequate fit $\Delta\chi^2_{(df=81)}=155.54, p < .001$, CFI = .95, TLI = .92, RMSEA = .08 (90% CI .06–.09), SRMR = .08. The Satorra Chi-square difference test between the restrictive model with equal factor loadings and the baseline model without indicates that the model without equal factor loadings fit significantly better, $\Delta\chi^2_{(df=4)}=73.64, p < .001$. We found that all loadings were equal except Total Number of Characters (TNC) between the two writing samples for grade 4. Turning to measurement invariance of intercepts, we found that the model without equal intercepts fit significantly better, $\Delta\chi^2_{(df=8)}=173.21, p = .001$. A follow-up analysis of

each intercept was conducted and the variables found to have equal intercepts were mean length of T-Unit, number of different characters, mechanical errors made for the alternative characters which have a similar orthographic form and the same pronunciation (i.e., MLT, NDW, ORE, and PHE), which suggested that the scales of these observed variables are the same for two writing samples for grade 4.

We examined the measurement invariance between writing sample A and writing sample B for grade 7. Similar to grade 4, two measures had correlated residuals across writing sample A and B, the Topic + Number of Key Elements ($r = .26, p = .001$), and Number of Different Characters ($r = .42, p < .001$). Results for tests of measurement invariance are presented in Table 5. The baseline model resulted in a good fit $\chi^2_{(df=77)}=99.83, p = .04, CFI = .98, TLI = .97, RMSEA = .04$ (90 % CI .01–.06), and SRMR = .05. The Satorra Chi-square difference test between the restrictive model with equal factor loadings and the baseline model without indicated that the model without equal factor loadings fit similar, $\Delta\chi^2_{(df=4)} p = .58$. Turning to measurement invariance for intercepts, we found that the model with equal intercepts fit more poorly, $\Delta\chi^2_{(df=8)}=22.29, p = .004$. Follow up analyses indicated that there were equal intercepts for all variables except Order and Number of Different Characters (i.e., NDC), which suggested that the scales of all the observed variables measured for grade 7, except for Order and NDC, were scaled similarly across the two writing samples.

We examined the measurement invariance between grades 4 and 7 on writing sample A and writing sample B respectively using multi-group CFA (see Table 6). Note that all five factors are included for examination. For writing sample A, the baseline model resulted with a good fit $\chi^2_{(df=54)}=95.15, p < .001, CFI = .97, TLI = .94, RMSEA = .07$ (90 % CI .04–.09), and SRMR = .04. The model with equal loadings resulted with a significantly poorer fit $\Delta\chi^2_{(df=5)}=71.05, p < .001$. We examined each variable individually, and found that MLT and NDW had different loadings. We further tested the invariance on intercepts of the remaining variables and found that Sentence Copying did not have equal intercepts.

For writing sample B, the baseline model resulted in a good fit $\chi^2_{(df=53)}=109.78, p < .001, CFI = .96, TLI = .92, RMSEA = .08$ (90 % CI .05–.10), and SRMR = .06. The model with equal loadings resulted in a similar fit, $\Delta\chi^2_{(df=5)}=6.21, p = .29$. We tested the invariance of intercepts and determined that Order and TNC did not have equal intercepts.

In summary, the purpose of the analyses just described was to determine whether measurement invariance (i.e., whether the factors were the same) across 4th and 7th grades and across the two writing samples was supported by the data. Having established at least partial measurement invariance, we were then able to compare factor correlations and factor means across grades.

Comparing correlations across grades—We compared the factor correlations across grades in the following way. We fixed variances to be equal on corresponding factors across grades and then imposed the constraint that one covariance coefficient at a time was equal.

The fit of these models was compared to the fit of models without this constraint using a Chi-square difference test. In these models, factor loadings and intercepts previously found to be equal across grades were kept equal so that the corresponding factors were comparable across grades. For writing sample A, we found that the following correlations were identical across grade ($ps > .08$): macro-organization with complexity, macro-organization with mechanical errors, complexity with productivity, complexity with handwriting fluency, productivity with spelling and punctuation, productivity with handwriting fluency, and spelling and punctuation with handwriting fluency. For writing sample B, we further tested each correlation and found that the following correlations were equal ($ps > .06$): macro-organization with mechanical errors, complexity with productivity.

Comparing latent means across grades—We compared latent means of the five factors on writing sample A across grades, and found that grade 7 had significantly higher means for complexity, productivity, and handwriting fluency, and significantly lower means for mechanical errors ($ps < .001$). There was no difference for macro-organization. For writing sample B, the mean comparison of the five factors across grades 4 and 7 yielded the same pattern of differences as writing sample A ($ps < .01$). In summary, the factor correlations, which describe the latent structure of written composition, were largely identical across grade and writing samples. The major differences between grades were in the means of the factors. Compared to 4th grade writers, 7th grade writers wrote more, wrote faster, wrote more complexly, and made fewer errors.

DISCUSSION

In the present study, we applied a five-factor model of writing that was developed from analyses of English writing samples to Chinese writing samples provided 4th and 7th grade students. Despite marked differences in the characteristics of the two writing systems, the confirmatory factor analysis results provide evidence that a five-factor model of English written composition generalizes to Chinese writing samples. These results suggest that much of what underlies individual and developmental differences in writing reflects deeper cognitive and linguistic factors as opposed to the more superficial differences in the writing systems.

By supporting a multi-factor view of writing, the results of these studies appear to conflict with both the Yan et al. (in press) analysis of Chinese writing samples and the Mehta et al. (2005) analyses of English writing samples, both of which supported a unidimensional or single factor model. However, we believe the models may be addressing different aspects of writing. One potential explanation for these differences that needs to be examined in future studies concerns the nature of the variables that were analyzed. For the present study and for Wagner et al., with the exception of a single variable that was a rating of the logical ordering of ideas, all other the variables were quantitative measures of things like number of T-units. For the Yan et al. and Mehta et al. studies, the variables were qualitative ratings of various aspects of the written compositions. The pattern of results across these four studies suggests that quality ratings and quantitative counts may be tapping important yet different aspects of writing.

Consistent with Yan et al. and Wagner et al., handwriting fluency is related to a variety of aspects of written composition. Whether handwriting fluency ought to be considered an integral aspect of a model of written composition as is the case for the five-factor model, or as a predictor of written composition as was the case for Yan et al. is an interesting question for future research. For the Yan et al. study, a large set of substantively important predictors was available for use in predicting the quality of the writing samples. In this context, it was informative to include handwriting fluency among other predictors of writing to determine whether it made an independent contribution to prediction. For the present study and Wagner et al. (2011), the initial conceptualization of the five-factor model of writing included handwriting fluency as an integral aspect of written composition and a comprehensive set of predictors of writing was not available. Under these circumstances, it seemed to make more sense to include it as a factor in the model rather than as a sole predictor.

Turning to developmental differences, once again the five-factor model provided the best fit to both grades examined, and provides support for the model when applied to writing samples obtained from first through seventh grades. Developmental differences are reflected primarily in differences in latent means of the factors as opposed to the factor structure itself.

Finally, the results suggest that a five-factor model of English written composition generalizes to multiple writing prompts although some parameters of the model may vary across writing samples.

Limitations and future research

Although coding variables in SALT is believed to be a strength of the present study and the previous study by Wagner et al., it will be important in future research to demonstrate that the fact that the five factor model of writing applies to both Chinese and English writing samples is not limited to the use of the SALT coding system. It could be the case that SALT codes relatively universal aspects of language, to the neglect of important language specific or written language specific elements of writing. A first step in addressing this potential limitation would be to develop other indicators of the factors of the five factor model that are not based on SALT codes.

A second limitation of the present study is that the design was cross-sectional rather than longitudinal. A longitudinal design might have provided more power to detect more subtle developmental differences in writing.

It also is important to acknowledge that our study only addressed a narrow aspect of the translation aspect of writing, and ignored important questions about how writing is related to both oral language and reading. We think it is important that future studies of the five-factor model of writing include measures of oral language and of reading to enable determination of what is specific to writing as opposed to general to reading or oral language.

Finally, it is important to follow up the results of correlational studies with intervention studies that attempt to manipulate performance on key constructs to better understand their interrelations (MacArthur et al., 2006).

Acknowledgments

This research was funded by NICHD Grant P50 HD052120 to Richard K. Wagner.

References

- Abbott RD, Berninger VW, Fayol M. Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. *Journal of Educational Psychology*. 2010; 102:281–298.10.1037/a0019318
- Alves, RA.; Castro, SL.; Sousa, L.; Stromqvist, S. Influence of typing skill on pause-execution cycles in written composition. In: Torrance, M.; van Waes, L.; Galbraith, D., editors. *Writing and cognition: Research and applications*. 2007. p. 55-65.
- Bentler PM. Comparative fit indexes in structural models. *Psychological Bulletin*. 1990; 107:238–246. [PubMed: 2320703]
- Bentler, PM. EQS structural equations program manual. Encino, CA: Multivariate Software; 1995.
- Bentler PM, Bonett DG. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*. 1980; 88:588–606.
- Berninger VW. Coordinating transcription and text generation in working memory during composing: Automatic and Constructive Process. *Learning Disability Quarterly*. 1999; 22:99–112.
- Berninger V. Development of language by hand and its connections to language by ear, mouth, and eye. *Topics of Language Disorders*. 2000; 20:65–84.
- Berninger V. Highlights of programmatic, interdisciplinary research on writing. *Learning Disabilities Research & Practice*. 2009; 24:69–80. [PubMed: 19644563]
- Berninger, V.; Amtmann, D. Preventing written expression disabilities through early and continuing assessment and intervention of handwriting and/or spelling problems: Research into practice. In: Swanson, HL.; Harris, K.; Graham, S., editors. *Handbook of learning disabilities*. New York: Guilford; 2003.
- Berninger V, Graham S. Language by hand: A synthesis of a decade of research on handwriting. *Handwriting Review*. 1998; 12:11–25.
- Berninger, VW.; Swanson, HL. Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. In: Butterfield, EC., editor. *Children's writing: Toward a process theory of the development of skilled writing*. Hampton Hill: JAI Press; 1994. p. 57-81.
- Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bolleny, KA.; Long, JS., editors. *Testing structural equation models*. Newbury Park: Sage; 1993. p. 136-162.
- Chanquoy L, Alamargot D. Working memory and writing: Evolution of models and assessment of research. *Annee Psychologique*. 2002; 102:363–398.
- Christensen CA. The role of orthographic-motor integration in the production of creative and well-structured written text for students in secondary school. *Educational Psychology*. 2005; 25:441–453.
- Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*. 2001; 108:204–256. [PubMed: 11212628]
- Connelly V, Campbell S, MacLean M, Barnes J. Contribution of lower-order letter and word fluency skills to written composition of college students with and without dyslexia. *Developmental Neuropsychology*. 2006; 29:175–198. [PubMed: 16390293]
- Connelly V, Dockrell J, Barnett J. The slow handwriting of undergraduate students constrains overall performance in exam essays. *Educational Psychology*. 2005; 25:99–107.
- Connelly V, Hurst G. The influence of handwriting fluency on writing quality in later primary and early secondary education. *Handwriting Today*. 2001; 2:50–57.

- Coulmas, F. The writing systems of the world. Oxford & New York: Basil Blackwell; 1991.
- De La Paz, S.; Graham, S. Dictation: Applications to writing for students with learning disabilities. In: Scruggs, T.; Mastropieri, M., editors. *Advances in learning and behavioral disorders*. Vol. 9. Greenwich, CT: JAI Press; 1995. p. 227-247.
- DeFrancis, J. The ideographic myth. In: Erbaugh, MS., editor. *Difficult characters: Interdisciplinary studies of Chinese and Japanese writing*. Columbus, OH: National East Asian Language Resource Center, Ohio State University; 2002. p. 1-20.
- Dockrell J, Lindsay G, Connelly V. The impact of specific language impairment on adolescents' written text. *Exceptional Children*. 2009; 75:427-436.
- Fayol, M.; Alamargot, D.; Berninger, V., editors. *Translation of thought to written text while composing: Advancing theory, knowledge, methods, and application*. New York: Psychology Press; (in press)
- Graham S. The role of production factors in learning disabled students' compositions. *Journal of Educational Psychology*. 1990; 82:781-791.
- Graham S, Berninger V, Abbott R, Abbott S, Whitaker D. The role of mechanics in composing of elementary school students: A new methodological approach. *Journal of Educational Psychology*. 1997; 89(1):170-182.
- Graham S, Harris KR. Almost 30 years of writing research: Making sense of it all with *The Wrath of Khan*. *Learning Disabilities Research & Practice*. 2009; 24:58-68.
- Greg, L.; Steinberg, R. *Cognitive processes in writing*. Hillsdale, NJ: Erlbaum; 1982.
- Grigorenko, EL.; Mambrino, E.; Priess, DD., editors. *Writing: A mosaic of new perspectives*. New York: Psychology Press; 2011.
- Guan CQ, Liu Y, Chan DHL, Perfetti CA. Writing strengthens orthography and alphabetic-coding strengthens phonology in learning to read Chinese. *Journal of Educational Psychology*. 2011; 103(3):509-522.
- Hayes, J. A new framework for understanding cognition and affect in writing. In: Levy, CM.; Ransdell, S., editors. *The science of writing*. Mahwah, NJ: Erlbaum; 1996. p. 1-27.
- Hayes, J.; Flower, L. Identifying the organization of writing processes. In: Gregg, LW.; Steinberg, ER., editors. *Cognitive processes in writing*. Hillsdale, NJ: Erlbaum; 1980. p. 3-30.
- Hoosain, R. *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum; 1991.
- Hooper SR, Costa LJC, McBee M, Anderson KL, Yerby DC. Concurrent and longitudinal neuropsychological contributors to written language expression in first and second grade students. *Reading and Writing*. 2011; 24:221-252.
- Hu L, Bentler PM. Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1-55.
- Hu L, Bentler PM, Kano Y. Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*. 1992; 112:351-362. [PubMed: 1454899]
- Ju D, Jackson NE. Graphic and phonological processing in Chinese character identification. *Journal of Reading Behavior*. 1995; 27:299-313.
- Kellog RT. Competition for working memory among writing processes. *The American Journal of Psychology*. 2001; 114:175-191. [PubMed: 11430147]
- Kellog RT. Working memory components in written sentence generation. *The American Journal of Psychology*. 2004; 117:341-361. [PubMed: 15457806]
- Kline, RB. *Principles and practice of structural equation modeling*. New York, NY: Guilford Press; 2011.
- LaBerge D, Samuels SJ. Toward a theory of automatic information processing. *Cognitive Psychology*. 1974; 6:283-323.
- Leck KJ, Weekes BS, Chen MJ. Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory and Cognition*. 1995; 23:468-476. [PubMed: 7666760]
- Levy, CM.; Ransdell, S., editors. *The science of writing: Theories, methods, individual differences, and applications*. Mahwah, NJ: Lawrence Erlbaum; 1996.

- Lin D, McBride-Chang C, Shu H, Zhang Y, Li H, Zhang J, et al. Small wins big: Analytic Pinyin skills promote Chinese word reading. *Psychological Science*. 2010; 21:1117–1122.10.1177/0956797610375447 [PubMed: 20581343]
- Lui H-M, Leung M-T, Law S-P, Fung RS-Y. A database for investigating the logographeme as a basic unit of writing Chinese. *International Journal of Speech-Language Pathology*. 2010; 12(1):8–18. doi:0.3109/17549500903203082. [PubMed: 20380245]
- MacArthur, CA.; Graham, S.; Fitzgerald, J., editors. *Handbook of writing research*. New York: Guilford Press; 2006. p. 275-290.
- MacCallum RC, Browne MW, Sugawara HM. Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*. 1996; 1:130–149.
- McCutchen D. A capacity theory of writing: working memory in composition. *Educational Psychology Review*. 1996; 8(3):299–325.
- McCutchen, D. Cognitive factors in the development of children's writing. In: MacArthur, CA.; Graham, S.; Fitzgerald, J., editors. *Handbook of writing research*. New York: Guilford; 2006. p. 115-130.
- Mehta PD, Foorman BR, Branum-Martin L, Taylor WP. Literacy as a unidimensional multilevel construct: Validation, sources of influence, and implications in a longitudinal study in grades 1 to 4. *Scientific Studies of Reading*. 2005; 9:85–116.
- Miller, J.; Chapman, R. *Systematic analysis of language transcripts (Version 7.0)* [computer software]. Madison, WI: Waisman Center, University of Wisconsin-Madison; 2001.
- Muthén, LK.; Muthén, BO. *Mplus user's guide*. 6. Los Angeles, CA: Muthén & Muthén; 1998–2010.
- Nelson, NW.; Bahr, C.; Van Meter, A. *The writing lab approach to language instruction and intervention*. Baltimore, MD: Paul H. Brookes; 2004.
- Nelson NW, Van Meter A. Assessing curriculum-based reading and writing samples. *Topics in Language Disorders*. 2002; 22:35–59.
- Nelson NW, Van Meter AM. Measuring written language ability in narrative samples. *Reading & Writing Quarterly*. 2007; 23:287–309.
- Olive T, Alves RA, Castro SL. Cognitive processes in writing during pauses and execution periods. *European Journal of Cognitive Psychology*. (in press).
- Olive T, Kellogg RT. Concurrent activation of high- and low-level production processes in written composition. *Memory & Cognition*. 2002; 30:594–600. [PubMed: 12184560]
- Perfetti, CA.; Guan, CQ. Effect of repeated writing practice. In: Guan, CQ., editor. *Written language studies across culture; Symposium conducted at the meeting of the American Educational Research Association Annual Meeting; Vancouver, Canada*. 2012 Apr.
- Perfetti CA, Zhang S. Very early phonological activation in Chinese reading. *Journal of Experimental Psychology: Learning Memory and Cognition*. 1995; 21(1):24–33.
- Peverly ST. The importance of handwriting speed in adult writing. *Developmental Neuropsychology*. 2006; 29:197–216. [PubMed: 16390294]
- Puranik C, Lombardino L, Altmann L. Assessing the microstructure of written language using a retelling paradigm. *American Journal of Speech-Language Pathology*. 2008; 17:107–120. [PubMed: 18448599]
- Rapp B, Benzing L, Caramazza A. The autonomy of lexical orthography. *Cognitive Neuropsychology*. 1997; 14:71–104.
- Satorra, A. Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In: Heijmans, RDH.; Pollock, DSG.; Satorra, A., editors. *Innovations in multivariate statistical analysis*. London: Kluwer; 2000. p. 233-247. A Festschrift for Heinz Neudecker
- Satorra A, Bentler PM. A scaled differences Chi-square test statistic for moment structure analysis. *Psychometrika*. 1988; 66(4):507–514.10.1007/BF02296192
- Scott C, Windsor J. General language performance measures in spoken and written discourse produced by school-age children with and without language learning disabilities. *Journal of Speech, Language, and Hearing Research*. 2000; 43:324–339.

- Shu, H.; Anderson, RC. Learning to read Chinese: The development of metalinguistic awareness. In: Wang, J.; Inhoff, AW.; Chen, H-C., editors. *Reading Chinese script: A cognitive analysis*. Mahwah, NJ: Lawrence Erlbaum; 1999. p. 1-18.
- Torrance, M.; Galbraith, D. The processing demands of writing. In: MacArthur, C.; Graham, S.; Fitzgerald, J., editors. *Handbook of writing research*. New York: Guilford; 2006. p. 67-80.
- Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3(1):4–69.
- Venesky, R. *The structure of English orthography*. The Hague, The Netherlands: Mouton; 1970.
- Venesky, R. *The American way of spelling*. New York: Guilford Press; 1999.
- Wagner RK, Puranik CS, Foorman B, Foster E, Wilson LG, Tschinkel E, et al. Modeling the development of written language. *Reading and Writing*. 2011; 24:203–220. [PubMed: 22228924]
- Weekes BS, Chen MJ, Lin YB. Differential effects of phonological priming on Chinese character recognition. *Reading and Writing: An Interdisciplinary Journal*. 1998; 10:201–222.
- Weekes B, Yin W, Su IF, Chen MJ. The cognitive neuropsychology of reading and writing in Chinese. *Language and Linguistics*. 2006; 7:595–617.
- Whitaker D, Berninger V, Johnston J, Swanson L. Intra-individual differences in levels of language in intermediate grade writers: Implications for the translating process. *Learning and Individual Differences*. 1994; 6:107–130.
- Yan CMW, McBride-Chang C, Wagner RK, Zhang J, Wong AMY, Shu H. Writing quality in Chinese children: Speed and fluency matter. *Reading and Writing: An Interdisciplinary Journal*. (in press).

Fig. 1.

Confirmatory factor analysis structure, standardized factor loadings, and inter-factor correlations of Passage A for Grade 4. † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Fig. 2.

Confirmatory factor analysis structure, standardized factor loadings, and inter-factor correlations of Passage B for Grade 4. † $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Fig. 3.

Confirmatory factor analysis structure, standardized factor loadings, and inter-factor correlations of Passage A for Grade 7. $^{\dagger}p < .10$; $*p < .05$; $**p < .01$; $***p < .001$

Fig. 4.

Confirmatory factor analysis structure, standardized factor loadings, and inter-factor correlations of Passage B for Grade 7. [†] $p < .10$; * $p < .05$; ** $p < .01$; *** $p < .001$

Table 1

Descriptive statistics for the composition and handwriting fluency variables of two writing samples of Grade 4 and Grade 7.

	Grade 4												Grade 7											
	Sample A						Sample B						Sample A						Sample B					
	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
<i>Macro-organization</i>																								
Topic	.97	.18	-5.40	27.53	.99	.11	-8.86	77.45	.92	.27	-3.08	7.56	.88	.33	-2.29	3.28								
Logical ordering or idea	2.09	.60	-.03	-.20	2.24	.60	-.14	-.48	2.10	.83	.06	-1.04	2.32	.94	-.02	-1.02								
Number of key elements	1.86	.52	-.17	.42	2.04	.54	.03	.53	1.91	.70	.12	-.95	2.05	.78	-.08	-1.35								
<i>Complexity</i>																								
Mean length of T-units	25.12	7.01	.96	1.34	22.98	9.00	2.19	7.95	32.16	12.32	2.88	15.76	30.53	11.41	1.15	2.29								
Clause density	13.07	3.24	2.42	10.35	10.46	2.27	.83	1.96	14.56	3.71	1	2.31	14.94	6.47	4.97	44.38								
<i>Productivity</i>																								
Total number of words	127.04	51.22	.29	-.65	103.54	46.70	.51	-.45	203.32	82.10	.20	-.40	196.60	81.42	.12	-.75								
# of different words	74.84	27.93	.77	.93	73.69	28.20	.20	-.73	145.91	59.93	.42	.21	146.13	56.66	.25	-.17								
<i>Spelling and punctuation</i>																								
# of phonological error	.66	1.18	2.14	4.27	.80	.93	.88	-.27	.41	.79	2.12	4.34	.38	.72	2.15	5.13								
# of orthographical errors	.70	.96	1.33	1.15	.60	1.05	2.33	6.21	.26	.59	2.68	7.90	.27	.70	3.56	14.60								
# of period errors	.92	1.87	3.18	11.98	.71	1.56	2.66	7.25	.00	.00	—	—	.01	.08	12.92	167.00								
<i>Handwriting fluency</i>																								
Stroke printing fluency	33.00	13.24	.59	.20	33.00	13.24	.59	.20	67.17	21.31	.88	1.23	67.17	21.43	.88	1.18								
Sentence copying fluency	14.26	4.02	.86	1.53	14.26	4.02	.86	1.53	30.44	8.54	2.29	9.44	30.44	8.54	2.29	9.44								

Table 2

Correlations between compositional and handwriting fluency variables for Grade 4.

	1	2	3	4	5	6	7	8	9	10	11	12
1 Topic	—	.33***	.30***	-.19*	.04	.03	.03	.04	-.09	-.07	.12	.21**
2 Logical ordering of ideas	.04	—	.73***	-.06	.11	.52***	.44***	.15	-.02	.04	.41***	.43***
3 Number of key elements	.22**	.76***	—	-.15*	.12	.48***	.44***	.19*	.02	-.02	.41***	.52***
4 Mean length of T-units	-.02	-.22**	-.18*	—	.28***	-.04	-.09	.05	.13	-.11	-.07	-.02
5 Clause density	.03	.16*	.07	.36***	—	.06	.00	.11	.22**	-.18*	-.02	.21**
6 Total number of words	.13	.68***	.50***	.03	.45***	—	.90***	.26**	.08	.09	.35***	.36***
7 # of different words	.15	.69***	.51***	.04	.45***	.96***	—	.23**	.07	.14	.25**	.34***
8 # of phonological error	-.02	.13	.07	.10	.13	.29***	.28***	—	.18*	.16*	.07	.10
9 # of orthographical errors	.06	.09	-.02	-.07	-.09	.07	.08	.19*	—	.10	.10	.08
10 # of period errors	.05	-.05	.00	.12	.12	.02	.02	-.07	-.02	—	-.06	-.01
11 Stroke printing fluency	-.12	.33***	.12	.02	.26**	.47***	.43***	.35***	.13	-.15	—	.44***
12 Sentence copying fluency	.12	.34***	.33***	-.05	.11	.39***	.39***	.05	-.05	-.04	.44***	—

N = 160. Sample A are in the upper diagonals. Sample B are in the lower diagonals.

*
 $p < .05$;**
 $p < .01$;***
 $p < .001$

Table 3

Correlations between compositional and handwriting fluency variables for Grade 4.

	1	2	3	4	5	6	7	8	9	10	11	12
1 Topic	—	.22**	.18*	-.01	-.01	-.22**	-.23**	.08	.02		.08	-.09
2 Logical ordering of ideas	.40***	—	.72***	-.19*	-.10	.46***	.39***	.14	.20**		.00	.00
3 Number of key elements	.42***	.82***	—	-.25**	-.19*	.49***	.44***	.08	.27***		-.02	.02
4 Mean length of T-units	-.01	-.14	-.17*	—	.47***	.00	.03	-.05	-.09		.05	-.01
5 Clause density	-.04	-.11	-.10	.43***	—	.06	.10	.06	-.13		-.01	-.07
6 Total number of words	.03	.53***	.47***	.22**	.10	—	.95***	.23**	.20**		-.03	.07
7 # of different words	.01	.51***	.47***	.19*	.13	.94***	—	.22**	.15		-.04	.05
8 # of phonological error	.03	.04	.09	-.05	-.20*	.04	.01	—	.24**		.07	.05
9 # of orthographical errors	-.06	.05	.01	-.06	-.13	-.01	-.04	.15*	—		-.02	.05
10 # of period errors	.03	.06	.00	.15*	.05	-.03	-.02	-.04	-.03		—	
11 Stroke printing fluency	.16*	.10	.10	-.12	.00	.05	.02	-.03	-.05	-.08	—	.56***
12 Sentence copying fluency	.20**	.09	.07	.00	.01	.09	.07	-.09	-.07	.03	.56***	—

N = 160. Sample A are in the upper diagonals, Sample B are in the lower diagonals.

*
 $p < .05$;**
 $p < .01$;***
 $p < .001$

Table 4

Model fit of five-factor CFA by sample and grade.

	Grade 4		Grade 7	
	Sample A	Sample B	Sample A	Sample B
Satorra-Bentler Scaled χ^2	88.81	81.39	34.20	3.81
<i>df</i>	36	35	28	28
<i>p</i> value	<.001	<.001	.19	.33
RMSEA (90% CI)	.09 (.07, .12)	.09 (.06, .11)	.04 (.00, .07)	.02 (.00, .06)
CFI	.92	.94	.99	.99
TLI	.87	.91	.98	.99
SRMR	.06	.07	.04	.05

CFI Comparative Fit Index, *TLI* Tucker Lewis coefficient; *RMSEA* root mean square error of approximation, *SRMR* standardized root mean squared residual

*
p < .05;

**
p < .01;

p < .001

Table 5

Examination of measurement invariance between samples A and B for Grades 3 and 7.

		<i>df</i>	χ^2	CFI	TLI	RMSEA (90% CI)	SRMR	χ^2	<i>df</i>
<i>Grade 4</i>									
Model 1	Baseline model	77	125.17***	.97	.95	.06 (.04–.08)	.07		
Model 2 (compared to Model 1)	Model with equal loadings	81	155.54***	.95	.92	.08 (.06–.09)	.08	73.64***	4
Model 3 (compared to Model 1)	Model with equal loadings except TNW	80	131.27***	.96	.95	.06 (.04–.08)	.07	6.58	3
Model 4 (compared to Model 3)	Model 3 + equal intercepts	88	33.28***	.83	.76	.13 (.12–.15)	.21	173.21***	8
Model 5 (compared to Model 3)	Model 3 + equal intercepts on MLT, NDW, ORE, PHE	84	139.17***	.96	.94	.06 (.05–.08)	.08	7.73	4
<i>Grade 7</i>									
Model 1	Baseline model	77	99.83*	.98	.97	.04 (.01–.06)	.05		
Model 2 (compared to Model 1)	Model with equal loadings	81	101.57	.98	.97	.04 (.00–.06)	.05	2.86	4
Model 3 (compared to Model 2)	Model 3 + equal intercepts	89	131.66**	.96	.95	.05 (.03–.07)	.05	22.29**	8
Model 4 (compared to Model 3)	Model 3 + equal intercepts except order and TNW	87	106.92	.98	.98	.04 (.01–.06)	.05	6.23	6

CFI/Comparative Fit Index, *TLI*/Tucker Lewis coefficient, *RMSEA* root mean square error of approximation, *SRMR* standardized root mean squared residual, *TNW* total number of words, *MLT* mean length of T-units, *NDW* number of different words, *ORE* number of orthographical errors, *PHE* number of phonological errors

* $p < .05$;

** $p < .01$;

*** $p < .001$

Table 6

Examination of measurement invariance between Grades 3 and 7.

		χ^2	df	CFI	TLI	RMSEA (90% CI)	SRMR	df	χ^2
<i>Sample A</i>									
Model 1	Baseline model	95.15***	54	.97	.94	.07 (.04–.09)	.04		
Model 2 (compared to Model 1)	Model with equal loadings	17.33***	59	.90	.85	.11 (.09–.12)	.09	5	71.05***
Model 3 (compared to Model 1)	Model with equal loadings except MLT and NDW	101.06***	57	.96	.94	.07 (.04–.09)	.05	3	5.92
Model 4 (compared to Model 3)	Model 3 + equal intercepts	114.18***	60	.95	.93	.07 (.05–.09)	.08	3	11.48**
Model 5 (compared to Model 3)	Model 3 + equal intercepts on MLT, NDW, and SENTENCE	102.21***	59	.96	.95	.06 (.04–.08)	.06	2	1.47
<i>Sample B</i>									
Model 1	Baseline model	109.78***	53	.96	.92	.08 (.05–.10)	.06		
Model 2 (compared to Model 1)	Model with equal loadings	115.28***	58	.95	.93	.08 (.06–.10)	.07	5	6.21
Model 3 (compared to Model 2)	Model 2 + equal intercepts	175.17***	63	.91	.87	.10 (.08–.12)	.08	5	52.06***
Model 4 (compared to Model 2)	Model 2 + equal intercepts except ORDER and TNW	12.11***	61	.95	.93	.08 (.06–.10)	.07	3	4.84

CFI/Comparative Fit Index, TLI/Tucker Lewis coefficient, RMSEA root mean square error of approximation, SRMR standardized root mean squared residual, TNW total number of words, MLT mean length of T-units, NDW number of different words, ORDER logical ordering of idea, SENTENCE sentence copying fluency

* $p < .05$;

** $p < .01$;

*** $p < .001$