Thinking aloud during idea generating and planning before written translation: Developmental changes from ages 10 to 12 in expressing and defending opinions

Matt Davidson1* and Virginia Berninger1

Abstract: This interdisciplinary research, drawing on cognitive psychology and linguistics, extended to middle childhood past research during early childhood or adulthood on thinking aloud prior to written composing. In year 5 of a longitudinal study of typical writing, when cohort 1 was in grade 5 (n = 110 ten year-olds) and cohort 2 in grade 7 (n = 97 twelve year-olds), a cross-sectional study was conducted. Children were first asked to think aloud while they generated ideas and second while they planned their essays to express and defend their opinions on a controversial topic in the region of the United States where they lived. Third, they wrote their essays. Their think-aloud protocols were audio-recorded and later transcribed into writing for analysis. The authors developed and applied rating scales for quality of idea generating and planning in the written transcriptions and quality of opinion expression, opinion defense, organization, and content in the essays children wrote after thinking aloud; total number of words in essays was also counted. Seventh graders scored significantly higher than fifth graders on quality of idea generation but not planning, and higher on all variables rated for quality in the written essays including length. Quality of expressing opinions and defending opinions were

ABOUT THE AUTHORS

Matt Davidson, MEd, University of Washington, undertook this research as a result of his experiences talking with writers he taught in university writing centers. His main research interests are related to writing and assessment. He currently teaches International Baccalaureate Psychology in Vietnam, and will continue studying the relationship between writing and speaking about writing as a PhD student.

Virginia Berninger, professor emeritus University of Washington, was the principal investigator of NICHD-funded research on writing in typical learners and those with specific learning disabilities from 1989 to 2008 and 2011 to 2016. She taught children to write in public schools, taught graduate students in school psychology how to assess writing and link it to instruction and undergraduates and graduate students about writing development from interdisciplinary perspectives, and conducted clinical assessment of writing in a range of clinical populations and consulted with schools regarding writing for over 35 years.

PUBLIC INTEREST STATEMENT

Twenty-first century social media offers many opportunities to express opinions. Yet, at school in many countries students must write essays both expressing one’s own opinion and others’ opinions and making and defending a case taking into account possible different perspectives. Also in democracies, social studies courses at school often prepare students for future voting by having them view debates in which different perspectives are expressed and defended. In this study of children in the United States at the end of elementary school (aged 10 on average) or in the middle of middle school (aged 12 on average), ability to express and defend opinions with awareness of others’ perspectives were separable processes and only began to become integrated in the 12-year olds. However, at each age level individual differences occurred. Application of results to writing instruction at school and to communication on social media is discussed.
uncorrelated in grade 5, but moderately correlated in grade 7. Whether idea generating or planning quality explained unique variance in essays varied with coded written essay variables and grade. Educational applications of results for assessment, assessment-instruction links, instruction in social studies, and theory of mind in persuasive essay writing are discussed.

Subjects: Development Studies, Environment, Social Work, Urban Studies; Communication Studies; Education

Keywords: opinion essays; argumentative essays; persuasive essays; idea generating; planning

1. Introduction
The current study is one of the last in a five-year, overlapping cohort longitudinal study of typical development of writing, reading, listening comprehension, and oral expression. Because participating children were assessed annually for half a day in the first two to four months of a grade, there was opportunity to collect a rich set of data across early to middle childhood and early adolescence regarding transcription modes (Alstad et al., 2015; Berninger, Abbott, Augsburger, & Garcia, 2009; Berninger et al., 2006), levels of language in language by hand (writing) and language by eye (reading)—word, sentence, and text (Abbott, Berninger, & Fayol, 2010; Berninger & Abbott, 2010; Niedo & Berninger, 2016), integrated reading-writing (Altemeier, Jones, Abbott, & Berninger, 2010; Berninger & Abbott, 2010; Niedo & Berninger, 2016), and cognitive processes in writing such as idea generation (Berninger, Richards, et al., 2009; Hayes & Berninger, 2010) and planning, reviewing, and revising (Berninger, Abbott, Whitaker, Sylvester, & Nolen, 1995; Berninger, Fuller, & Whitaker, 1996; Berninger, Whitaker, Feng, Swanson, & Abbott, 1996). These studies employed nationally standardized tests with norms and experimenter-designed writing tasks with research norms or researcher-designed coding schemes.

During the final year when cohort 1 was in fifth grade and cohort 2 was in seventh grade, a cross-sectional study was conducted related to the research team’s evolving research on the role of idea generation and planning in translation of cognitions into oral and written language (Fayol, Alamargot, & Berninger, 2012; Hayes, 1996). Idea generating was of interest for understanding the nature of the content of what was being translated and whether the process might be characterized as flow (Kellogg, 1994; Kellogg & Whiteford, 2012). Planning was of interest for understanding the role of strategies in organizing the content that was being translated (Hayes, 1996). During planning, the writer thinks about how to organize generated ideas for translation of those ideas into language (Hayes, 2012). Think-alouds capture the idea generating and planning processes as translated into oral language before being translated into written language during written composing. Idea generating and planning were studied in separate think-alouds to evaluate if the observed pre-writing, thinking processes in oral think-alouds for idea generation and planning did indeed differ from one another.

Idea generating and planning were studied for a single genre—argumentative (also referred to as persuasive)—for four reasons. First, a number of recent studies have shown that translation outcomes vary across genres. Olinghouse, Santangelo, and Wilson (2012) examined the validity of “single-occasion, single-genre, holistically scored” pieces of writing (p. 55). Correlations across pairs of different genres (narrative, informative essay, and persuasive essays) showed only moderate correlations between genre pairs for the holistic quality ratings (values 0.37–0.48) and lower correlations for genre elements (values 0.12–0.23) in fifth graders.

Second, argumentative/persuasive composing requires expression of opinion/s and defense of opinion/s. For recent review of research on argumentative genre, see Ferretti and Fan (2016). As such this genre of writing has an important contribution to educating school age children and youth for their future participation in democracy if they live in a country where the people have a voice in
their government (Parker, 2005, 2010). Although oral discussion can be an effective pedagogical practice for helping students learn to express their opinions and defend them to others (Parker, 2010), especially if encouraged to argue for both sides of issue being debated, so can writing to express and defend their opinions before or after such oral discussion. Writing a persuasive/argumentative essay has the potential advantage of encouraging every student to engage at the same time in expression and defense of one’s own perspective and then after follow-up oral discussion in which they learn of others’ perspectives, consider others’ perspectives in their next argumentative/persuasive essay on the same topic. This approach could lead to both greater engagement of the audience than mere observation of a debate and greater awareness of the perspectives of others in developing civic literacy foundations for democracy (cf., Parker, 2005, 2010).

Third, theory of mind (Frith & Frith, 2010) predicts that learning to understand the self’s own perspective is an important steppingstone to understanding the perspectives of others, which may differ from one’s own perspective. Indeed theory of mind is beginning to be recognized as one of the potential contributions to and beneficiaries of written argumentative/persuasive writing (Panthee & Williamson, 2014). That is, expression and defense of opinions, whether written or oral or both, affords an important opportunity to study the dynamic interplay of the self and of other in learning to compose in writing, which may be crucial to developing the sense of self that underlies learning to self-regulate both learning and behavior (Frith & Frith, 2010). Indeed, writing research has shown the effectiveness of teaching explicit strategies for self-regulating the composing process (see MacArthur, Graham, & Fitzgerald, 2016). At the same time, writing is uniquely sensitive to “other” in that much of what is written is for an audience of others. On the one hand, a genre that involves expressing one’s own opinion or perspective is ideally suited for helping the writer develop a sense of self, which then can play a role in learning to self-regulate the expression of one’s own ideas. On the other hand, a genre that requires defending one’s own opinions or perspectives and constructing arguments to persuade others requires the transformation of those ideas so that others can be persuaded to one’s own view/s. That is, the writer has to go beyond knowledge telling (expressing) to knowledge transforming (making and defending the case) (Bereiter & Scardamalia, 1987).

Fourth, learning to express and defend opinions in writing has relevance beyond language arts. Across the curriculum, students are often given assignments that require articulating a perspective about what they read, learning from teacher or classmate presentations, or in class discussions and defending one’s own position and/or persuading others to adopt that perspective too. Thus, in the current study, the topic for expressing and defending an opinion was relevant to content areas of the curriculum in both science (mountains in nature) and social studies (controversial social justice issue in the geographical region where students lived, regarding whether volcanic glaciers well known in the area should be named for the much earlier indigenous migrants from Asia and South America or the more recent explorers from Europe).

The rationale for the use in the current study of a think-aloud protocol, rather than written production of idea generation or planning as is often used in writing research, was threefold. First, cognitive research has shown that the process of transforming mental representations into written language is supported by working memory (Kellogg, 1996; Kellogg, Whiteford, Turner, Cahill, & Mertens, 2013; Swanson & Berninger, 1996). One component of working memory in Baddeley’s model of working memory is the phonological loop (Baddeley, 1986, 2007), which may support idea generating, planning, and translating processes, either covertly (one’s inside voice) or overtly (thinking aloud). Thus, it was of interest to study the potential contribution of the overt phonological loop while developing writers generated ideas or planned strategies orally. Second, research on planning during the upper elementary and middle school grades has often focused on the use of written outlines (Koutsotaas & Gray, 2013) or other visual schemas. Past research with middle childhood and early adolescent writers found that written planning was often confounded with the translation process itself (Berninger et al., 1996). Oral think-alouds might avoid this confounding and capture pre-planning before the written translation begins. Third, there is research evidence for the instructional benefits of thinking aloud during the translation processes during composing in early childhood (see...
2. Research aims, tested hypotheses, and methodological approach

The first research aim was therefore to audio record students thinking aloud as they generated ideas and planned their essays to express and defend their opinions on a controversial topic. The second research aim was to develop and apply a scheme for rating the quality of idea generating and quality of planning observed in the think-alouds transcribed into writing by the research team. The third research aim was to develop and apply a scheme regarding the quality of opinions expressed, quality of opinions defended, quality of content, and quality of organization in the written essays after the oral think-alouds and also count the number of words as an index of essay length.

Three hypotheses were tested. The first hypothesis was that quality ratings and length would show developmental change in that the seventh graders would have higher ratings than the fifth graders. The second hypothesis was that correlations among the ratings for idea generating, planning, and translating would be significant but of relatively low magnitude. That is, because the variables rated for degrees of quality represent separable processes that are not totally redundant with each other. The third tested hypothesis was that the quality ratings for idea generating and planning in the think-alouds would explain unique variance in the ratings for the essay quality variables or essay length. However, these relationships would vary according to the ratings for specific cognitive processes in the think-alouds or written language variables in essays entered into each of the multiple linear regressions at each grade level.

The methodological approach was interdisciplinary. It was informed by the cognitive processes of writing (Fayol et al., 2012; Hayes, 1996, 2012; Kellogg, 1994) and the over two decades of experience (1989–2016) of the research team in linguistic coding of levels of language during translation (e.g. Niedo & Berninger, 2016) and quality of composing in written narratives or essays on many outcomes variables in cross-sectional as well as longitudinal and intervention studies (e.g. Berninger, 2009).

3. Method

3.1. Participants

All aspects of the conducted research were in compliance with the approved institutional review board (IRB) protocol at the university where it was carried out and the ethical guidelines of the American Psychological Association. Students were recruited for the longitudinal study from a school district near the university where the research was conducted via a letter sent by the school district to parents of kindergarteners and second graders. Interested parents contacted the research team who explained the study, conducted a phone interview to verify that the student was probably a typically developing writer without developmental or specific learning disabilities or other disorders, and obtained parental informed consent and child assent, using procedures that had been approved by the IRB for human participants in research.

Then an appointment was scheduled for an assessment at the university during autumn of first grade (cohort 1) or third grade (cohort 2) for participants who attended 51 different schools in the very large urban school district near the university in the Pacific Rim of the US. This appointment was the first annual half-day assessment during the second to fourth month of the school year for the five-year longitudinal study in which the same children were reassessed annually from first to fifth grade (cohort 1) or third to seventh grade (cohort 2) with cohorts overlapping in grades 3–5. Parents received an honorarium to defray travel expenses and an annual comprehensive report with their child’s assessment results, but the children and their parents were not paid to participate in the study.
The sample represented the ethnic diversity in the area: European American (64.8%), Asian American (23.4%), African-American (6.3%), Hispanic (1.6%), Native American (1.6%), and other (2.3%) in fifth grade; and European American (65.5%), Asian American (21.2%), African-American (9.7%), Hispanic (0.9%), and other (2.7%) in seventh grade. Mothers (M) and fathers (F) varied in their level of education, ranging from not completing high school (cohort 1 M 7%, F 12.5%; cohort 2 M 7.1%, F 7.1%) to high school or more than high school but less than college degree (cohort 1 M 7.0%, F 12.5%; cohort 2 M 11.5%, F 14.2%) to college degree (cohort 1 M 45.3%, F 39.8%; cohort 2 M 50.4%, F 36.3%) or graduate degree (cohort 1 M 45.3%, F 39.8%; cohort 2 M 30.1%, F 35.4%).

Attrition rate was low (cohort 1, n = 128 in year 1 and n = 114 in year 5; cohort 2, n = 113 in year 1 and n = 99 in year 5); the reasons for the 14 in each cohort discontinuing participation across the five years was mainly due to moving and not being able to come to the university. Prior to the annual Literacy Trek session at the university, the research team planned jointly how to make it an engaging, enjoyable, and educational experience for participants that year. During the course of the study that year, the team discussed at weekly group meetings students assessed that week to monitor what went well, any problems observed, and if so how to avoid them in the future. The annual time at the university was presented to the students as an opportunity to be a college student for a day, in contrast to most students having to graduate from high school before they even become a college student. Also the activities they engaged in were varied and changed often to maintain interest and included both tasks like a regular day at school and special activities students normally do not get to do at school. To maintain energy levels and attention, nutrition breaks, and movement breaks were interspersed throughout the 4 h morning or 4 h afternoon assessment session with one graduate research assistant who performed individual assessment and showed an interest in how the current school year was going for each student. See Berninger and Hayes (2012) for how comprehensive the assessment was for 20 representative participants across each of the five years.

Only fifth graders in cohort 1 and seventh graders in cohort 2 who had usable data for both the think-alouds and the written essays (all but 4 in cohort 1 and all but 2 in cohort 2) were included in the data analyses in the current study. (Some data were lost because of initial technology issues with the audio recording.) The sample analyzed included the year 5 fifth graders (n = 110; 44% male, 56% female, mean age was 128.50 months, SD = 3.64) and seventh graders (n = 97; 49.5% male, 50.5% female; mean age was 151.21 months, SD = 3.71).

3.2. Procedures

To minimize that results were solely a result of differing background knowledge on the topic of the argumentative/persuasive essay, prior to the think-alouds, content was provided for background knowledge by the examiner reading a text orally as the child read it silently, and by providing photographs of both mountains and other content in the background material. Children were then asked to think aloud for five minutes to generate ideas for an essay on the topic of whether original Native American names of the mountains or names given to the mountains by European explorers should be used for two well-known snow-capped mountains in the region, which had been formed from volcanoes. They were then asked to think aloud for five minutes to make a plan how to organize the expression of their opinions and a defense of their opinions about this topic. Specific instructions were as follows, in this order.

1. “Read along silently while I read aloud a text about the controversies about two mountains, both of which have experienced volcanic eruptions. Controversies mean that different people have different opinions or points of view.”

2. “Now, I want you to write an essay in which you explain the different points of view about each controversy, give your opinion or point of view about each controversy, and defend your opinion or point of view and try to convince the reader against the opposing opinion or point of view. First, I want you to brainstorm your ideas. You don’t have to write yet. Just let your idea generator pump out ideas. I will tape record them so I can remember them.”
(3) After 5 min, the examiner continued with directions: “Now, I want you to plan how you will go about writing your essay. What are your goals? How will you organize the essay? How will you start? How will you end? You’re thinking out loud. So I can remember them, again I will tape record you.”

(4) After 5 min, the examiner continued with these directions: “Now, you can write your essay titled, Defending My Opinions on Some Controversies about Mt. St. Helens and Mt. Rainier. You will have 5 min to write it.”

Each think-aloud (for idea generating and for planning) was audiotaped for 5 min. If the student ceased thinking aloud, the examiner could prompt up to twice to keep thinking. Total time in minutes spent thinking aloud, if less than the full five minutes for each think-aloud, was recorded. Examiners on the research team who performed the individual assessments transcribed the audiotaped think-alouds they collected into writing for subsequent analyses.

3.2.1. Constructing the coding schemes and evaluating inter-rater reliability

Initially, the first author examined a sub-sample of 10 fifth grade and 10 seventh grade transcribed think-alouds and developed schemes for rating quality of idea generating, and planning during think-alouds and for indicators of quality of opinion expression and defense of opinion and content and organization in written essays: number of words was used as an indicator of length. See Appendix A for the schemes that were developed and could be used by other researchers. Then the first and the second author jointly inspected the transcripts of think-alouds and the written essays to determine whether these schemes accounted for all that was observed in the transcriptions and essays and made a few modifications to ensure that they did. Next, the first author and second author used the same schemes to independently rate another set of think-alouds and written essays for a second subset of 10 participants (5 male, 5 female) from each grade level. Then they met again to compare their ratings. For grade 5, agreement for rating idea generating was 100%, and for rating planning was 80%. For grade 7, agreement for rating idea generating and rating planning was 80%. Following these same procedures, the two raters achieved an overall rate of 89% agreement (grade 5) and 88% agreement (grade 7) on the rated variables in the written essays. Where the two independent raters had rated differently, the differences were discussed and readily resolved; and minor adjustments were made in the wording of the scheme to facilitate reliability in future. Finally, the first author coded the remaining transcribed think-alouds and written essays and rated the remaining participants on the items in the schemes, but consulted occasionally with the second author for cases that were ambiguous or more than one code in the scheme might apply. These rare instances were discussed and resolved.

4. Results

Because total time engaged in thinking aloud was not related to any of the other measures it was not considered further in the analyses.

4.1. First tested hypothesis—Developmental change in quality ratings and essay length

Grade 7 participants scored significantly higher than grade 5 participants on the quality ratings for express opinion, defend opinion, organization quality, content quality, and length (number of words) in written essays. However, grade 7 participants were rated higher than the grade 5 participants only on quality of idea generating—not quality of planning. See Table 1 for the ANOVA results and means and standard deviations.

4.2. Second tested hypothesis—Inter-relationships among coded variables

Table 2 summarizes the correlations among all the ratings or essay length for both grades 5 and 7. In grade 5, all the quality ratings for essays or essay length were significantly correlated with each other, but at modest magnitudes. Express Opinion and Defend Opinion exhibited the correlation of the highest magnitude, but only explained about 50% of the variance in each other. Of interest, neither of the quality ratings for the cognitive variables (idea generating and planning) was
correlated with each other in grade 5; and only idea generating quality was significantly correlated with each of the quality ratings or length for written essays. Planning quality was significantly correlated only with quality ratings for content and organization. See Table 2.

In grade 7, the results for the correlations were often the same as in grade 5. All the written essay variables were significantly correlated with each other, but at a low magnitude. In contrast to grade 5, however, quality ratings for idea generating and planning were now significantly correlated, but at a low magnitude. Also, in grade 7, idea generating quality was significantly correlated at a low magnitude with all written essay quality ratings but not length; however, planning quality was correlated only with essay length, content quality, and organization quality. See Table 2.

### Table 1. Significant mean differences in grades 5 and 7 for written language variables in writing products and cognitive processes

| Measure          | Grade 5 | Grade 7 | F(1, 204) |
|------------------|---------|---------|-----------|
|                  | M       | SD      | M         | SD        |           |
|                   | n = 110 |         | n = 97    |           |           |
| **Writing products** |         |         |           |           |           |
| Opinion          | 3.19    | 1.24    | 3.98      | 1.13      | 23.01***  |
| Defense          | 2.54    | 1.13    | 3.97      | 1.05      | 85.59***  |
| Organization     | 3.08    | 1.04    | 3.90      | 0.90      | 54.11***  |
| Content          | 2.97    | 0.88    | 3.91      | 0.94      | 69.28***  |
| Words written    | 55.19   | 23.25   | 82.27     | 25.04     | 37.91***  |
| **Cognitive processes** |         |         |           |           |           |
| Idea generating  | 2.45    | 0.77    | 2.75      | 0.52      | 10.56***  |
| Planning         | 2.13    | 0.90    | 2.26      | 0.77      | 1.39      |

***p ≤ 0.001.

### Table 2. Correlations among quality ratings for idea generating and planning, genre-specific explaining and defending opinions, content and organization quality, and length (words)

|          | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------|---|---|---|---|---|---|---|
| **Grade 5** |   | r | p | r | p | r | p |
| 1. Opinion | - |  |   |  |   |  |  |
| 2. Defense | 0.76*** | - |  |   |  |  |  |
| 3. Content | 0.57*** | 0.66*** | - |  |   |  |  |
| 4. Org | 0.56*** | 0.59*** | 0.71*** | - |  |  |  |
| 5. Words | 0.35** | 0.39*** | 0.50*** | 0.51*** | - |  |  |
| 6. Idea gen | 0.31** | 0.23* | 0.21* | 0.25* | 0.23* | - |  |
| 7. Planning | 0.10 | 0.02 | 0.20 | * | 0.32** | 0.40 | 0.21 |

|          | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|----------| --- | r | p | r | p | r | p |
| **Grade 7** |   |   |   | r | p | r | p |
| 1. Opinion | - |   |   |   |   |   |   |
| 2. Defense | 0.67*** | - |   |   |   |   |   |
| 3. Content | 0.54*** | 0.66*** | - |   |   |   |   |
| 4. Org | 0.49*** | 0.62*** | 0.66*** | - |   |   |   |
| 5. Words | 0.37*** | 0.57*** | 0.59*** | 0.49*** | - |   |   |
| 6. Idea gen | 0.26* | 0.22* | 0.25* | 0.26* | 0.08* | - |   |
| 7. Planning | 0.18 | 0.15 | 0.40*** | 0.24* | 0.45*** | 0.37*** | - |
4.3. **Third tested hypothesis: Predicting written language translation from idea generating and planning**

4.3.1. **Predicting opinion expression**
In both grade 5 and grade 7, the multiple linear regression with quality of idea generating and planning as predictors accounted for significant shared variance in the written essay variables, but only quality of idea generating explained unique variance in grades 5 and 7. See Table 3.

4.3.2. **Predicting defense of opinion**
In grade 5, the multiple linear regression with quality of idea generating and planning as predictors accounted for significant shared variance in the written essay variables, but only idea generating quality in grade 5 explained unique variance. In grade 7, the predictors did not account for significant shared variance. See Table 4.

4.3.3. **Predicting organization quality**
At both grade levels, the multiple linear regression with quality of idea generating and planning as predictors explained significant shared variance in the written essay variables. However, quality of idea generating and planning explained unique variance only in grade 5, and neither of the predictors did in grade 7. See Table 5.

4.3.4. **Predicting content quality**
At both grade levels, the multiple linear regression with quality of idea generating and planning as predictors explained significant shared variance in the written essay variables, but only in grade 7 did quality of planning explain unique variance. See Table 6.

4.3.5. **Predicting length (number of words)**
At both grade levels, the multiple regression with quality of idea generating and planning as predictors explained significant shared variance in the written essay variable, but only quality of planning explained unique variance in the written essay variable at both grades 5 and 7. See Table 7.

5. **Discussion**

5.1. **Tested hypotheses**

5.1.1. **First tested hypothesis**
The first hypothesis was mostly supported. Quality of idea generating while thinking aloud and quality of opinion expression, opinion defense, content, and organization and the length of written essays improved significantly from grades 5 to 7. Only quality of planning while thinking aloud did not improve from grades 5 to 7. This pattern of results is consistent with idea generating and planning

| Table 3. Multiple regressions for predicting expression of opinion quality from cognitive processes in grades 5 and 7 |
|----------------------------------------------------------|
| **F(df)** | Adjusted $R^2$ | $\beta$ | t | $p$ |
|---------------------------|------------------|--------|---|---|
| Grade 5                   | $F(2, 106) = 5.63^{**}$ | 0.079  |    |    |
| Idea generating           |                   | 0.301  | 3.19 | ** |
| Planning                  |                   | 0.036  | 0.38 | ns |
| Grade 7                   | $F(2, 94) = 3.73^*$  | 0.054  |    |    |
| Idea generating           |                   | 0.223  | 2.09 | *  |
| Planning                  |                   | 0.092  | 0.86 | ns |

*p < 0.05.
**p < 0.01.
***p < 0.001.
being separable processes. The items identified for the rating scale for quality of idea generation showed that individuals differed not only in the number of ideas that flowed (Kellogg, 1994) but also they differed in meta-cognitive awareness of whether the ideas were or were not relevant to opinion expression or opinion defense (see items that were rated for quality in scheme in Appendix A).

5.1.2. Second tested hypothesis

The second hypothesis was supported in that the written essay (translation product variables) and the cognitive variables for idea generating and planning were correlated; but the magnitude of the correlations tended to be low or modest at best. Thus, these are separable, but inter-related cognitive processes during translation into oral language prior to translation into the written language.

| Table 4. Multiple regressions for predicting defense of opinion quality from cognitive processes in grades 5 and 7 |
|---------------------------------------------------------|
| **F(df)** | Adjusted $R^2$ | $\beta$ | $t$ | $p$ |
|-----------|-----------------|--------|-----|-----|
| **Grade 5** | | | | |
| Idea generating | $F(2, 106) = 3.0^*$ | 0.036 | 0.236 | 2.45 | * |
| Planning | | | −0.029 | −0.297 | ns |
| **Grade 7** | | | | |
| Idea generating | $F(2, 94) = 2.61$ n.s. | 0.032 | 0.184 | 1.703 | ns |
| Planning | | | 0.085 | 0.786 | ns |

*p < 0.05.

**p < 0.01.

***p < 0.001.

| Table 5. Multiple regressions for predicting organization quality from cognitive processes in grades 5 and 7 |
|---------------------------------------------------------|
| **F(df)** | Adjusted $R^2$ | $\beta$ | $t$ | $p$ |
|-----------|-----------------|--------|-----|-----|
| **Grade 5** | | | | |
| Idea generating | $F(2, 106) = 7.95$*** | 0.114 | 0.187 | 2.02 | * |
| Planning | | | 0.274 | 2.96 | ** |
| **Grade 7** | | | | |
| Idea generating | $F(2, 94) = 4.61$** | 0.070 | 0.197 | 1.86 | ns |
| Planning | | | 0.163 | 1.54 | ns |

*p < 0.05.

**p < 0.01.

***p < 0.001.

| Table 6. Multiple regressions for predicting content quality from cognitive processes in grades 5 and 7 |
|---------------------------------------------------------|
| **F(df)** | Adjusted $R^2$ | $\beta$ | $t$ | $p$ |
|-----------|-----------------|--------|-----|-----|
| **Grade 5** | | | | |
| Idea generating | $F(2, 106) = 3.95$* | 0.052 | 0.173 | 1.81 | ns |
| Planning | | | 0.167 | 1.74 | ns |
| **Grade 7** | | | | |
| Idea generating | $F(2, 94) = 9.60$*** | 0.152 | 0.122 | 1.20 | ns |
| Planning | | | 0.351 | 3.47 | *** |

*p < 0.05.

**p < 0.01.

***p < 0.001.
variables in the translation products. These results are consistent with the separable cognitive processes in the Hayes (1996, 2012) models.

5.1.3. Third tested hypothesis
The third hypothesis was supported. On the one hand, the cognitive predictor variables expressed orally contributed via shared variance to all the written language translation variables in written essays, except for defending opinion in grade 7. However, as predicted, which cognitive predictors explained unique variance in the written language translation variables varied both between grades 5 and 7 and across the coded written language translation variables. In both grades 5 and 7, only idea generating explained unique variance in expressing one’s opinion, but only planning explained unique variance in length. Only at grade 5 did idea generating uniquely explain defending one’s own opinion. Only at grade 5 did idea generating and planning uniquely explain organization quality. Only at grade 7 did planning uniquely explain content quality.

5.2. Educational applications—Psychoeducational assessment
The results of the current study should be generalized only to one genre, namely argumentative/persuasive essay writing, and two genre-specific features—expressing and defending opinions. Nevertheless, the results illustrate the importance of assessment of writing skill that is genre-specific (Olinghouse et al., 2012). Not only does assessment of writing ability in developing writers require multiple writing samples (Olinghouse et al., 2012), but also normed standardized measures of translation products may not capture all the relevant cognitive processes that contribute to written essay writing such as idea generating and planning (Hayes, 1996, 2012; Kellogg, 1994; Kellogg & Whiteford, 2012; Kellogg et al., 2013).

Of interest, the “I don’t know” response occurred during think-alouds for both idea generating and planning; it was only coded if nothing else was produced during the oral think-aloud for idea generating or for planning. Such “I don’t know” responses also occur during individual administration of normed measures. Future research might explore the motivational, affective, executive function, and metacognitive variables that may contribute to over use of this response across a variety of kinds of assessment tools in some individual students. In the current study, the frequency of these responses was relatively low.

5.3. Educational applications for assessment-instruction links
The current study extends past research showing that thinking aloud prior to and during writing is an evidence-based instructional strategy for composing during early childhood (Berninger, 2009; Berninger & Chanquoy, 2012) as well as a useful research and assessment tool in adults (Chenoweth & Hayes, 2003). As reviewed in Berninger and Chanquoy, a kindergarten teacher who taught the “What I Think I Can Say, What I Say I Can Write Study” was successful in teaching kindergartners in a school with low income, racial minority students to write and even read proficiently for grade level

| Table 7. Multiple regressions for predicting coded number of words in essay from cognitive process in grades 5 and 7 |
|---------------------------------|-------------|--------|-----|
| | F(df) | Adjusted $R^2$ | $\beta$ | $t$ | $p$ |
|---------------------------------|-------------|--------|-----|
| Grade 5                         | F(2, 106) = 10.96*** | 0.156 |     |     |
| Idea generating                 |             | 0.127 | 1.41 | ns |
| Planning                        |             | 0.369 | 4.09 | ***|
| Grade 7                         | F(2, 94) = 12.78*** | 0.197 |     |     |
| Idea generating                 |             | -0.104 | -1.06 | ns |
| Planning                        |             | 0.491 | 4.99 | ***|
| **p < 0.05.**                   |             |        |     |     |
| **p < 0.01.**                   |             |        |     |     |
| ***p < 0.001.**                 |             |        |     |     |
because they read to each other what they wrote. This think-aloud strategy was also used in subsequent randomized control writing instruction studies for first, second, and third graders who were low achievers in handwriting or spelling to transfer taught transcription skills to translation during composing and shown to be effective (for review, see Berninger, 2009). As shown in the current study, thinking aloud prior to written translation may be helpful for some writers during middle childhood and early adolescence as well when writing in the genre of argumentative/persuasive essays. Such thinking aloud has been used successfully by the first author in a college writing center with adult English Language Learners (ELLs), many of whom are immigrants (unpublished data), but further research is needed on this approach.

Although visual aids are also helpful in generating ideas and planning strategies for writing about them (e.g. Koutsoftas & Gray, 2013), talking out loud also seems to facilitate some writers’ ability to translate their thinking into oral language first, and then from oral language into written language. The current study does not make a case for always or only using oral think-alouds in teaching writing, but does provide support for thinking aloud orally as being one tool for writing assessment with potential applications to instruction. Future research might address how oral think-alouds might most effectively be combined with written planning strategies for teaching composing of specific genres at specific grade levels or used alone for some writing goals. In sum, additional research is needed on the potential benefits of thinking aloud with and without visual aids before composing with upper elementary, secondary, and postsecondary students, both monolinguals and multi-linguals, and with and without specific learning disabilities in writing, for a variety of writing genres and academic writing assignments.

In addition, the lack of developmental improvement in quality of planning (see first tested hypothesis) suggests that explicit instruction in self-regulated strategies for planning is especially needed during this developmental period from grade 5 in upper elementary school to grade 7 in middle school (see MacArthur et al., 2016). It may take longer to master planning for the genre of argumentative/persuasive essay writing than other genres. However, an alternative explanation is that planning for this genre might benefit from a combination of oral thinking aloud and written outlines or diagrams rather than oral think-alouds alone. Yet, another possible explanation is that planning continues throughout translation and is not adequately captured only prior to translating; or translating may begin during planning. That is, it may be easier to separate processes of idea generating and written translating than of planning and written translating. Future research might explore whether these alternative explanations have merit.

Also, the results for the second tested hypothesis support the contribution of oral idea generating and planning before written argumentative/persuasive essay writing, but suggest that both individual differences and developmental changes may be operating. Formative assessment may provide clues for helping classroom teachers meet individual writers’ instructional needs (see Bennett, 2010; Harlen & James, 1997). Also see MacArthur, Graham, and Fitzgerald (2013) for specific evidence-based, effective writing instructional strategies linked to cognitive and written language translation skills.

**5.4. Educational applications for writing instruction in social studies**

Individual fifth and seventh graders varied in the quality of their opinion expression and opinion defense, which suggests that there are individual differences in developing writers learning to articulate their own perspectives and learning to understand and address the perspectives of others. Ongoing writing instruction in the argumentative/persuasive genre across the curriculum, coupled with oral thinking aloud and oral discussion with peers, may facilitate more students learning to both express their own opinions and defend them along with learning to understand the perspectives of others. Over time, this integrated written and oral language approach across the curriculum can prepare developing writers for future participation as citizens in democratic governments. See Parker (2005, 2010).
5.5. Educational applications for theory of mind

Many different social variables are relevant to learning to write. Writing is typically aimed at an audience of others, is often co-constructed with others who provide feedback to use in revising, and shared with others by authors orally reading what they have written. However, expressing opinions and creating arguments to defend those opinions are particularly sensitive social cognition variables. Persuading others requires sensitivity to the perspectives of others. Anticipating others’ thinking and social intents and motivations, which is key to developing a sense of self and related self-regulation skills, that is, theory of mind (Frith & Frith, 2010), is beginning to be applied to research on adult writers (Panthee & Williamson, 2014). However, theory of mind could have many applications to assessing and teaching writing earlier in schooling, as the current study illustrates. Argumentative/persuasive writing is an ideal genre for investigating theory of mind—awareness of others’ views and perspectives, especially when such views and perspectives may differ between writer—the self—and the audience—others. Building a case to persuade others to adopt one’s own perspectives requires awareness of both self and others. Such an approach to instruction might provide a valuable alternative or supplement to social media, which tends to facilitate expression of one’s own opinion or perspective but necessarily taking into account the perspectives of others in doing so.

5.6. Limitations and future research directions

The findings of the current study are appropriately generalized only to one country—the United States—and its primary language—English and educational policies (see Common Core Collaboration, 2009a, 2009b). However, similar studies are being conducted in other countries where not only languages but also educational practices and policies differ among each other as well as from those in the current study. At some point, there will probably be a sufficient body of evidence to support generalizations based on cross-country comparisons and multiple genres about which aspects of writing development transcend country and language. See Schoonen (2005, 2012) and van den Bergh, De Maeyer, van Wiejen, and Tillema (2012) for further discussion of the issues related to appropriate generalization of writing research findings across languages and countries.

Future research might be directed toward the alternative explanations that were proposed earlier for the results for the first and second tested hypotheses and implications of the results of the third tested hypothesis for developmental and individual differences in the relationships of cognitive and written language translation product variables. Future research should also address whether the development of the writer’s sense of self that develops through expressing and defending one’s own opinions in argumentative/persuasive genre writing might play a broader role in learning to write and applying self-regulated strategies to writing.

6. Conclusions

Expressing and defending opinions are genre-specific features of essay writing for making a persuasive argument that develop over time during middle childhood, but are probably not mastered until later in one’s writing development. The current assessment results show that ability to express one’s own opinion in persuasive essays is separable from ability to defend one’s own opinions taking into account the perspectives of others; but these separable writing skills become more integrated in 12-year olds than 10-year olds. Future instructional research for idea generation and planning for persuasive essay writing may identify effective ways for teaching developing writers to become self-regulated writers who can communicate across their own perspective (the self of the author) and the perspectives of others (the audience), who may or may not share the same opinions and perspectives as the author. Teaching developing writers to both express and defend opinions politely may also prepare them for becoming future citizens in a democratic government who are open to and grasp the perspectives of others as well as their own and for being more thoughtful, pro-social participants in social media.
Abbott, R., Berninger, V., & Fayol, M. (2010). Longitudinal opinions, Matt Davidson & Virginia Berninger, Cogent Psychology (2016), 3: 1276514.

References
Abbott, R., Berninger, V., & Fayol, M. (2010). Longitudinal relationships of levels of language in writing and between writing and reading in grades 1 to 7. Journal of Educational Psychology, 102, 281–298. http://dx.doi.org/10.1037/a0019318
Althard, Z., Sanders, E., Abbott, R., Barnett, A., & Hendersen, S. (2009). fMRI activation related to nature of ideas generated and differences between good and poor writers during idea generation. British Journal of Educational Psychology, 79, 77–93. http://dx.doi.org/10.1348/978185409421949
Berninger, V., Whitaker, D., Feng, Y., Swanson, H. L., & Abbott, R. (1996). Assessment of planning, translating, and revising in junior high writers. Journal of School Psychology, 34, 23–52. http://dx.doi.org/10.1016/0022-4053(95)00024-0
Chenoweth, N. A., & Hayes, J. R. (2003). The inner voice in writing. Written Communication, 20, 99–118. http://dx.doi.org/10.1177/0741088303253572
Common Core Collaboration. (2009). Common Core Collaboration among state leaders, including governors and state commissioners of education from 48 states, two territories and the District of Columbia. Common core standards for English language arts. Retrieved from www.corestandards.org/ELA-Literacy/W/5/ Common Core Collaboration. (2009). Common Core Collaboration among state leaders, including governors and state commissioners of education from 48 states, two territories and the District of Columbia. Common core standards for literacy in history/social studies, science, and technical subjects. Retrieved from www.corestandards.org/ELA-Literacy/WB/
Fayol, M., Almargot, D., & Berninger, V. (Eds.). (2012). Translation of thought to written text while composing: Advancing theory, knowledge, methods, and applications. New York, NY: Guilford. Frith, U., & Frith, C. (2010). The social brain: Allowing humans to boldly go where no other species has been. Philosophical Transactions of the Royal Society B: Biological Sciences, 365, 165–176. doi:10.1098/rstb.2009.0160
Horlen, W., & Bynum, J. (2012). Assessment and learning: Differences and relationships between formative and summative assessment. Assessment in Education: Principles Policy & Practices, 4, 365–379. doi:10.1007/s11886-012-0240-5

Funding
This research was supported by HD50HO071764 and HD25858 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) at the National Institutes of Health (NIH), USA.

Competing Interests
The author declares no competing interest.

Author details
Matt Davidson1
E-mail: mattjd@uw.edu
Virginia Berninger1
E-mail: vwb@uw.washington.edu
1 University of Washington, Seattle, WA, USA.

Citation information
Cite this article as: Thinking aloud during idea generating and planning before written translation: Developmental changes from ages 10 to 12 in expressing and defending opinions. Matt Davidson & Virginia Berninger, Cogent Psychology (2016), 3: 1276514.
Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences and applications (pp. 1–28). New York, NY: Routledge.

Hayes, J. R. (2012). Modeling and remodeling writing. Written Communication, 29, 369–388. doi:10.1177/0741088312451260

Hayes, J. R., & Berninger, V. (2010). Relationships between idea generation and transcription: How act of writing shapes what children write. In C. Brazerman, R. Krut, K. Lunsford, S. McLeod, S. Null, P. Rogers, & A. Stansell (Eds.), Traditions of writing research (pp. 166–180). New York, NY: Taylor and Frances/Routledge.

Kellogg, R. T. (1994). The psychology of writing. Oxford: Oxford University Press.

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy & S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences and applications (pp. 57–72). New York, NY: Routledge.

Kellogg, R. T., & Whiteford, A. P. (2012). The development of writing expertise. In E. L. Grigorenko, E. Mambrino, & D. D. Preiss (Eds.), Writing: A mosaic of new perspectives (pp. 109–124), New York, NY: Psychology Press.

Kellogg, R., Whiteford, A. P., Turner, C. E., Cahill, M., & Mertens, A. (2013). Working memory in written composition: An evaluation of the 1996 Model. Journal of Writing Research, 159–190. http://dx.doi.org/10.17239/jowr-2013.05.02.1

Koutsos, A. D., & Gray, S. (2013). A structural equation model of the writing process in typically-developing sixth grade children. Reading and Writing, 26, 941–966. doi:10.1007/s11145-012-9399-y

MacArthur, C., Graham, S., & Fitzgerald, J. (Eds.). (2013). Best practices in writing instruction (2nd ed.). New York, NY: Guilford.

MacArthur, C., Graham, S., & Fitzgerald, J. (Eds.). (2016). Handbook of writing research. New York, NY: Guilford.

Niedo, J., & Berninger, V. W. (2016). Strategies typically developing writers use for translating thought into the next sentence and evolving text: Implications for assessment and instruction. Open Journal of Modern Linguistics, 6, 276–292. doi:10.4236/ojml.2016.64029NIH

Appendix A

Coding schemes for oral think-alouds that were transcribed into writing

Think-alouds were coded on a scale of 0 to 3 using these schemes to rate quality along a scale ranging from a low of 0 to a high of 3.

Idea generating

(0) ‘I don’t know’.
(1) One or fewer ideas expressed.
(2) Discussion of ideas, but no support for the opinions expressed.
(3) Discussion of ideas as well as support/reasoning for opinions.

Planning

(0) ‘I don’t know’.
(1) Listing of content possibilities only.
(2) Organizational sequencing and content possibilities.
(3) Discussion of goals and/or strategies in addition to above.
Analyzing written essay product

Genre-specific elements of argumentative/persuasive written essays. Two scales, each ranging from a low 1 to high of 5, were used to code quality of expression of an opinion and quality of defense of the opinion. Higher ratings are given for essays that used phrases suited to the genre, as shown in the descriptions. The scale for opinion expression reflects whether the writer’s opinion on the controversy was clearly stated, for example, by use of phrases like “I think that” or “I believe” or “______ should”. The scale for defense of opinion reflects whether the text makes a reasoned defense of the writer’s view. Words like “because” or chains of reasoning are used.

Opinion ratings

(1) No inclusion of opinion.
(2) Opinion seems present, but not explicit.
(3) Some background about the controversy but vague expression.
(4) Explicit statement of opinion linked to background on controversy.
(5) Clearly explains the controversy and opinion.

Defense ratings

(1) No defense for opinion is provided.
(2) Defense for opinion is basic, such as “because they were here first.”
(3) Defense is more than basic such as a single clause as in (2).
(4) Defense includes more than one reason.
(5) Text makes a case for the opinion with clear evidence and/or reasons that are connected.

Organizational and content quality. A scale of 1–5 was used, as is described next, with 1 being less skilled and 5 being the most skilled.

Organization rating

(1) Ideas are presented in a list and are confusing.
(2) Ideas are in sentences, but do not progress logically.
(3) Organization is logical, but little to no framing (e.g. using topic sentences).
(4) Some framing of ideas, as well as logical progression.
(5) A lot of good framing as well as logical progression.

Content quality

(1) Content is not relevant.
(2) Some content is relevant but simply repeats facts that were given.
(3) All content is relevant, but simply repeats facts that were given.
(4) Content is relevant and elaborates on given facts.
(5) Content is relevant and sophisticated, modifies/uses the information.

Length. For each essay, the number of complete words was counted. Misspelled words but neither incomplete nor scratched out words were counted in the totals for number of words.
