Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.6, No.1, 2015

Effects of an Elementary Language Arts Unit on Critical Thinking, Reading, and Writing

Luke Duesbery[1*] Paul Justice[2]
1.Teacher Education, San Diego State University, 5500 Campanile Drive, San Diego, California, 92182-1153
2.Linguistics, San Diego State University, 5500 Campanile Drive, San Diego, California, 92182-7727
* E-mail of the corresponding author: duesbery@mail.sdsu.edu

Abstract
Teaching young students to think critically has always been important, however, as the United States transitions to a national set of learning standards which emphasizes higher-order thinking, it becomes essential. In this quasi-experimental study we evaluate the effects of exposure to the Journeys and Destinations (J&D) unit from the William and Mary language arts curriculum on students' critical thinking, reading, and writing in general education classrooms. The unit uses advanced-reading-level literature to teach the concept of change, critical reasoning, and advanced language arts skills. Students from nine fourth and fifth grade classrooms participated in the study; three used the William and Mary language arts model, while six were asked to use their normal language arts approach. At the beginning and end of the semester students were assessed with the Bracken Test of Critical Thinking, a test of syntactic reading fluency, and a curriculum-based measure of writing. Classroom observations were used to monitor the approaches being used and assess fidelity of implementation. Results indicated that students exposed to the J&D unit grew significantly in the area of critical thinking, while the comparison group did not. Both groups grew significantly in reading, and neither group experienced gains in writing. Results are discussed in the context of professional development needs, and the move to a national curriculum focused on developing critical thinking skills in all students.
Keywords: Curriculum, Elementary, Journeys and Destinations, Project Athena, Critical Thinking, Language Arts, Reading, Writing, Gifted

1. Introduction

The newly adopted Common Core State Standards (CCSS) were developed by the Council of Chief State School Officers (CCSSO) and the National Governors Association Center for Best Practices in the United States (2010). Since being released in 2010 these CCSSs in English Language Arts have been adopted by an overwhelming majority of states (ASCD, 2012). The objective of the standards is to provide a well-articulated set of expectations in core academic content areas. These expectations represent what the field thinks students will need to know before entering post-secondary education or the workforce.

With these new Standards, there is a transition in the United States away from *a mile wide and an inch deep*, toward helping students think and evaluate information critically - a move toward teaching students to think, not simply know. The Common Core State Standards Initiative Standards-Setting Criteria (2012, available at http://www.corestandards.org/resources) emphasizes a new focus on high-level content and the application of higher order thinking skills. They reported the standards include "high-level cognitive demands by asking students to demonstrate deep conceptual understanding through the application of content knowledge and skills to new situations". This emphasis should cause educators to take pause and re-evaluate the way they teach.

1.1 Purpose of the Study

The notion of teaching students to think critically is certainly not new. Elements of reasoning are documented and present in many curricula (Chandler, 2004). However it has been reported that there is a paucity of good research evaluating the efficacy of curricula at the elementary levels that focus on higher-order thinking (Van Tassel Baska, Bracken, Feng, & Brown, 2006). The current study responds to that need, and asks the broad research question: Can an instructional program that places emphasis on critical thinking within a content area like language arts lead to growth in related critical thinking skills? In general terms, this study investigates the impact of exposure to the William and Mary language arts unit, Journeys and Destinations (J&D), on critical thinking, reading, and writing skills among 4th and 5th grade students in heterogeneous classrooms. Specifically this study asks and answers the following questions:

1. Did students exposed to the J&D unit show greater gains on measures of critical thinking, reading, and writing than students who received the regular District curriculum?
2. Did teachers trained to implement the J&D unit do so with fidelity?
3. Did student engagement systematically differ in intervention and comparison classrooms?

1.2 College of William and Mary Curriculum

Originally called Project Athena, the William and Mary language arts units are based on the Integrated

Curriculum Model (ICM) (Van Tassel-Baska, 1986) and teach the concept of change, the reasoning process, advanced skills in literary analysis, linguistic competency, oral communication, and persuasive writing. The ICM focuses on content at a conceptual level with an emphasis on asking students to engage in critical thinking with advanced content (Van Tassel-Baska, 2003).

The series of nine language arts units are published by the Center for Gifted Education out of the College of William and Mary, and span content appropriate for grades two through eleven. In each of the units, students are asked to read passages or novels and then discuss and write about what they have read. During these activities, which are a mix of whole class, small group, and independent study, students learn new vocabulary, refine their persuasive writing and literary analysis skills, and practice communicating orally and thinking critically. The ICM calls upon teachers to question students with respect to the conceptual focus in each unit, to encourage them to think more deeply, and make connections between discrete pieces of literature.

The unit studied here, *Journeys and Destinations*, calls on students to examine and discuss the concept of change from different perspectives: relative to time, as random or orderly, and as positive or negative. Like the other units in the William and Mary series, *J&D* focuses on critical thinking, literary analysis, grammar, and persuasive writing. The 27 lessons emphasize the concept of change found in the written word, through a novel and several short stories and poems. Seven teaching strategies are integrated through the unit lessons.

1. The TABA Model of Concept Development: A strategy that asks students to generalize to a concept based on existing knowledge through inductive reasoning.
2. The Literature Web Model: A strategy that asks student to consider reading material from five perspectives – keywords, feeling, images/symbols, ideas, and writing structure.
3. The Vocabulary Web Model: A strategy that asks students focus in depth on interesting words.
4. The Hamburger Model for Persuasive Writing: A strategy that asks students to use the hamburger metaphor – with two levels of complexity – to organize their own writing.
5. The Reasoning Model: An integrated model based on Paul's Elements of Reasoning (1992) integrated through all lessons in the unit. These include considering assumptions, concepts, evidence, implications or consequences, inferences, issue, point of view, and purpose.
6. The Writing Process Model: A strategy that asks students to engage in a systematic process of writing, including prewriting, drafting, revising, editing, and sharing.
7. The Research Model: A strategy for independent or small group exploration that asks students to follow eight steps from the identification of an issue, to gathering and manipulating information, to making generalizations, and finally communicating findings.

In addition, emphasis is placed on asking higher-order thinking questions, the use of graphic organizers, and the inclusion of content from a multi-cultural perspective.

The importance of developing critical thinking skills in all students is of paramount importance, evidenced by the increasing emphasis on critical thinking in the new Common Core State Standards (Elliot, 2014). Scores on international tests demonstrate that American students, when compared to their international peers, are falling behind not only in math and reading, but similarly in both critical thinking and creativity (TIMSS, 2007) – those skills most important in an increasingly technologically dependent society. Units like the ones used in this study have the potential to close this gap, not only for high-achieving students, but also for those that are struggling with academics. By developing reasoning skills in young students, we increase student access to all content. Critical thinking helps students to answer questions they may have, but more than this, it helps students ask them.

1.3 Evidence of Effectiveness – High Ability Student

Early studies pointed to the efficacy of the William and Mary units with high ability students, with medium effect sizes for literary analysis and interpretation, and high effect sizes in persuasive writing (Feng, Van Tassel-Baska, Quek, Bair, & O'Neill, 2005; Van Tassel Baska, Johnson, Hughes, and Boyce, 1996; Van Tassel-Baska, Zuo, Avery, & Little, 2002). Several studies have also demonstrated the units' effectiveness in raising critical thinking, reading comprehension, and persuasive writing (Bracken, Van Tassel-Baska, Brown, & Feng, 2007; Van Tassel-Baska, Bracken, Feng, & Brown, 2008) as well as literary analysis. Most recently, Van Tassel-Baska and colleagues (2008) provided evidence of gains over time.

1.4 Evidence of Effectiveness – Teacher Factors

In the 2008 study, Van Tassel-Baska, Bracken, Feng, & Brown also examined teachers' instructional behaviour over 3 years in Title I schools. They concluded that two years of professional development and sustained support for implementation is needed to see instructional changes and fidelity of use, both related to growth in critical thinking and language arts achievement. Experimental teachers received higher ratings than control teachers on the Classroom Observation Scales (COS-R). However, students demonstrated higher levels of engagement on the COS-R in classrooms of veteran (3rd and 2nd year) as compared with newly trained experimental teachers.

Experimental teachers in this study received 4 days of training including 1 day of follow-up each year and implemented the curriculum over 12 to 16 weeks. This finding is in keeping with an early study (VanTassel-Baska, Johnson, Hughes, & Boyce, 1996) which indicated that the degree to which teachers accept the curriculum can moderate how much and how long they use the material, and the degree to which student are challenged when it is used.

1.5 Evidence of Effectiveness – Diverse Learners

The potential gains in critical thinking and content knowledge from using *J&D* for all students is of importance. A factor sustaining the achievement gap appears to be lower expectations for students from culturally and linguistically diverse and low-income backgrounds (McKown & Weinstein, 2008) manifested in instructional activities with less rigor and challenge (Rubie-Davies, 2007). Although developed for advanced learners, students with learning disabilities and typical learners as well as gifted students also make significant gains in critical thinking with these units (Hughes, 2000), and research has demonstrated positive results for heterogeneous groups of students from low-income backgrounds (Swanson, 2006).

2. Method

2.1 Teacher Participants

Twelve licensed teachers were recruited from a graduate level course in gifted education. Six of these teachers were randomly selected to receive a two-day training in the implementation of the J&D unit, led by the lead author of the unit. All teachers were female, and ranged in level of experience from three to more than twenty years of teaching. At the conclusion of the semester, only three teachers elected to remain in the intervention group, while all six remained in the comparison group. Teacher attrition, reported by those who did not complete the intervention, was attributed to the competing demands of the District required Units of Inquiry curriculum. The District Units of Inquiry are a series of lessons tightly aligned with State standards of learning, and while not scripted, are proscriptive in nature. Many administrators in the District require their teachers to follow the Units of Inquiry closely, which prevents many from adopting alternate units like the one used here.

2.2 Student Participants

While all students received pre- and post-tests, only data from those with parental consent were used during analyses. This yielded a comparison group of 87 and an intervention group of 41 students, spanning grades four and five. There is no reason to believe lack of parental support was systematic in either group – instead it is likely that parents simply forgot to return the permission forms. In both groups, demographics were similar; overall, approximately 6% of students were eligible for special education services, and 16% were identified as having intellectual gifts. Students in the sample mirrored district demographic proportions: approximately 46% Hispanic, 24% White, 12% African-American, 6% Filipino, and 8% Indo-Chinese or Asian. 30% of students were English Learners, and 59% were eligible for free or reduced meals. The question of differential functioning within groups was not the focus of this study, and as such characteristics of individual students were not coded for data analysis.

2.3 Procedure

Teachers in the intervention group received two full days of training in using the J&D unit. When teachers were ready to begin using the unit, the research team concurrently administered pre-tests in the areas of critical thinking, reading, and writing to both intervention and comparison groups. During the period in which teachers used the unit (approximately two months) teams of two highly experienced teachers observed each classroom, and observations were recorded for later analysis. At the end of the intervention period, a series of comparable post-tests was concurrently administered to both intervention and comparison groups. To minimize form effect, test forms were randomly counterbalanced across classrooms within groups.

2.4 Instruments

Five instruments were administered during the study. Teacher behaviour was documented with the Fidelity Checklist published within the *J&D* unit. Student behaviour was documented with the Classroom Observations Scales: Revised (COS-R, VanTassel-Baska, Avery, Struck, Feng, Bracken, Drummond, & Stambaugh, 2003), The Bracken Test of Critical Thinking (Bracken, Bai, Fithian, Lamprecht, Little, & Quek, 2003a), The Test of Syntactic Reading Fluency (Shinn, Deno, & Espin, 2000), and a curriculum based measure of writing (Shinn, 1989, p. 240).

Five experienced teacher volunteers conducted classroom observations for both the Fidelity of Implementation measure and the COS-R. Observers were trained over a two-day period, for a total of approximately six hours. During training, observers first discussed each item on each of the two observation instruments in an attempt to maximize consistency of understanding. Observers then practiced with several

video recordings of classrooms. After each video practice, observation data was collected and analyzed. Where marked disagreement occurred, discussion ensued. The process was repeated until the observation team was satisfied with the level of agreement. A final practice video observation was used to calculate inter-rater agreement for each possible pair of observers. Average inter-rater agreement across all pairs of observers was high for both the COS-R (88%) and the Fidelity Checklist (91%). Internal consistency of the measure has been documented in the range of r = .91-.93 (VanTassel-Baska, 2003)). Subsequent data analysis from field observations revealed similar levels of agreement, however, scores used for analysis were derived from consensus. The lead researcher and author, who is highly trained and experienced in educational measurement, administered the three academic achievement instruments in critical thinking, reading, and writing.

2.5 Teacher Behaviour: Fidelity of Implementation

The Fidelity Checklist asked observers to document the presence, and degree of implementation, of 12 instructional strategies present in the Journeys and Destinations unit. The checklist was originally designed by the College of William and Mary's curriculum design team for program evaluation purposes. Although it was unlikely we would see all components of the unit in such a short period of time, we considered it reasonable to expect to see at least some of the unit in the intervention group; likely more than in the comparison group. The three teachers in the intervention group reported covering between 90 and 100% of the unit content. Two experienced teachers administered the observation independently, and then prepared a third and final document through discussion and consensus. In general, the three teachers who received training in the intervention tended to implement more of the strategies associated with the unit. Marked differences occurred in the areas of oral literary analysis and structured questions. Fidelity data is presented in Table 1. Overall, the three teachers who received the VTB training tended to implement more of the strategies associated with the curriculum. In the areas of teaching grammar and encouraging reasoning the comparison teachers outscored the intervention teachers, by 0.3 and 0.5 respectively. In all other areas the intervention teachers outscored the comparison teachers, by as much as 1.5 in oral communications and 1.3 in literary analysis.

2.6 Student Behaviour: Classroom Observation Scales Revised

We assessed student behaviour during the study with the Classroom Observation Scales – Revised (COS-R) (VanTassel-Baska, Quek, & Xuemei Feng, 2007; VanTassel-Baska, et al., 2003). The COS-R is designed to determine the degree to which students exhibit behaviours associated with best practice in gifted and regular education, and is divided into nine board categories: general classroom behaviours, diverse self-selected or self-paced activities, formal problem solving strategies, critical thinking strategies, analysis and synthesis, transformative creative thinking strategies, explicit research strategies, dealing with content in depth, and dealing with multi-cultural content. During the study each classroom was observed for thirty minutes by two independent observers and each of the 25 behaviours were rated on a scale that asked observers to estimate the percent of students engaged in the behaviour – N/A, none, less than 25%, 25-50%, 51-75%, and >75%. These values were subsequently converted to a six-point scale where >75% was five, and N/A was coded zero. On occasion,a single observer was used when scheduling did not allow for the usual two. The consensus form, when available, was used to calculate within-group descriptive statistics.

2.7 Student Behaviour: Critical Thinking

The Bracken Test of Critical Thinking (TCT; Bracken et al., 2003) is a 45-item instrument designed to assess critical thinking skills in grades three through five. The TCT is based on Paul's model of critical thinking and includes the models' eight dimensions of thought: issue, purpose, concept, point of view, assumptions, evidence/information, inferences, and implications/consequences. The test is group-administered and consists of ten short scenarios relevant to the lives of elementary students. Each scenario is followed by a series of multiple-choice questions. Internal consistency of the measures used in this study has been documented in the range of r = .83 to .85 (Bracken, Bai, Fithian, Lamprecht, Little, & Quek, 2003b, p.24) for the grades tested. Over our pre- and post-tests, Cronbach's internal consistency reliability was reasonable (α = 0.77). To reduce fatigue, the test was administered over two days, approximately 30 minutes per session.

2.8 Student Behaviour: Reading Comprehension and Writing

The Test of Syntactic Reading Fluency (TSRF) is a group-administered measure in which students are asked to select from a list of appropriate words to complete sentences. Students are first delivered a standardized instruction protocol and a brief practice session, and are then given 180 seconds to read, quietly, from a single page of text. The first sentence is left intact, and after that, each seventh word is replaced by three words from which the student must choose the most appropriate. During silent reading, the administrator monitored the students to make sure they were actively reading and circling responses. After three minutes, the students were told to stop reading. A reading comprehension score is derived from the total words correctly selected less the

errors. Alternate forms reliability for the measures used in this study has been documented in the range of r = 0.81 (Shinn, Deno, & Espin, 2000).

The test of writing fluency is a brief and simple group measure that is used to estimate a student's writing skill. The student is presented with a starter prompt, is asked to first practice for two minutes, and then write for three minutes. The total number of words written is used as a proxy for level of writing skill (Deno, Marston, & Mirkin, 1982). Writing fluency tests have demonstrated technical adequacy with alternate forms reliability in the range of r = 0.76 to 0. 95, and inter-rater reliability in the range of r = 0.90 to 0.99 (McMaster & Espin, 2007).

2.9 Analyses

A combination of one-way repeated measures ANOVAs and descriptive statistics were used to compare group performance on the three measures. The -subjects factor was treatment condition, and the three tests (writing, reading, and critical thinking) served as dependent variables. All analyses were modeled with SPSS version 22.

3. Results

Results that follow are organized such that student observation data from the COS-R is presented first, followed by student achievement data in reading, writing, and critical thinking. Scores from classroom observation of students, in most cases, are the mean score of two independent observers. These behaviours are those thought to be consistent with how scholars think high ability students should be learning. Across all domains, observers used a five point scale, where four represents the behaviour occurred more than 75% of the time, three represents more than 50% of the time, two represents more than 25% of the time, and 1 represents less than 25% of the time. A score of zero meant the behaviour was not observed or there was no opportunity for the behaviour to be observed (i.e., it was not relevant to the lesson).

In all seven domains in which behaviours occurred, the intervention group outscored the comparison group. In general, students in the intervention group tended to display a higher frequency of general classroom behaviours associated with higher order thinking; behaviours like appearing challenged, applying learning, appearing thoughtful and evaluating evidence. The intervention group scored, on average, 3.5 on general classroom behaviours. In contrast, the comparison group scored, on average, 2.4. The largest difference occurred in the area of critical thinking, in which the intervention group scored, on average, 3.2 compared to just 0.9 in the comparison group. The complete series of summary statistics are provided in Table 2. In general, outside of the domains in which there was little variability to begin with, variance tended to remain fairly consistent between groups. In the domains of formal problem solving and creative thinking, no observed behaviours occurred in either group, and as a result are not reported in the table.

3.1 Student Achievement

Given the small sample size, there was insufficient power to perform a repeated measures ANOVA test to compare group differences, however, a paired samples t-test examining within-group differences could be used to test for statistical significance.

In the area of critical thinking, a significant within-group difference was found in the intervention group which grew by a average of 1.3 points on the Bracken Test of Critical Thinking, $t(40) = -2.371$, $p = 0.02$. In contrast, the comparison group growth of 0.7 points was not statistically significant, $t(86)=-1.833$, $p = 0.70$.

Both groups scored an average of about 15.5 before the six-week period on the reading test, and experienced statistically significant within-group growth. After the six weeks, the intervention group grew to 21.6, $t(40)=-7.16$, $p < 0.01$, and the control group grew to 21.9, $t(86)=-8.285$, $p < 0.01$.. This growth of about a point per week is generally thought of as acceptable reading growth (Hosp, Hosp, & Howell, 2007, p.47).

In terms of number of words written, there were no statistically significant changes in mean score; the intervention group wrote an average of about 47 words before the J&D unit was implemented, and fell by about 3 words over the course of the intervention. In contrast, the comparison group grew from about 46 to 47 words written. Descriptive statistics and effect sizes for all dependent variables are summarized in Table 3.

4. Discussion

The results found here are encouraging, but mixed. This, albeit small, sample of students who were exposed to the *J&D* unit experienced appreciable (d = 0.21) and statistically significant gains in the area of critical thinking. In contrast, the comparison group experienced no statistically significant gains. This effect size is small but in keeping with previous studies (Feng, Van Tassel-Baska, Quek, Bair, & O'Neill, 2005; Van Tassel Baska, Johnson, Hughes, and Boyce, 1996; Van Tassel-Baska, Zuo, Avery, & Little, 2002) and supports the value of the instructional unit for students of diverse ability levels. Given the randomization of classes to condition, it might be postulated that growth might have been due to the intervention.

The largest effect sizes were found in the area of reading; however, no differential effect between

groups was found. In essence, everyone, regardless of condition, was reading better at the end of the study.

On the measure of writing fluency, while no statistically significant changes occurred, the results are nonetheless worth discussion. The intervention group experienced negative growth. After the intervention they wrote fewer words (about one word per minute less) while the comparison group remained fairly stable. One of the teachers in the intervention group theorized that this confusing finding occurred because the metric was ill-suited to the outcome. The lead author of the unit affirmed this idea (Van Tassel-Baska, personal communication). In the unit, students learn to be better persuasive writers. This, in the end, may have caused them to write slower and better.

Not surprisingly, the students in the intervention group exhibited more of the behaviours associated with all of the activities associated with the J&D unit. This difference was most pronounced in the areas of critical thinking where the intervention group scored 2.3 points higher on the six-point scale, and analysis and synthesis in which the intervention group scored 1.6 points higher.

In terms of fidelity of implementation, the three teachers in the intervention group exhibited more of the teaching behaviours associated with the unit in almost all areas, indicating the intervention was delivered with a high degree of fidelity. It is worth noting that three of the original teachers dropped out of the study because of time constraints imposed by the competing District units. This may indicate a high level of support is needed to implement the unit.

In terms of direction for future research, efforts might be made to implement the teaching strategies explored here with a curriculum solely focused on Common Core State Standards. Understanding contextual mediating factors would be the next natural step. In addition, replicating this research on a larger sample of students would help to support findings.

The findings from this study are encouraging. The two-month language arts unit led to small but appreciable gains in critical thinking, a skill that is gaining importance with the adoption of the Common Core Standards. The strategies used in the unit can be adopted by teachers and integrated with their current practice. If implemented with fidelity, it would seem likely they might also lead to similar gains.

References

Association for Supervision and Curriculum Development (Nov. 9, 2014). *Status of State Adoption of Common Core State Standards*. Retrieved from http://www.ascd.org/common-core-state-standards/common-core-state-standards-adoption-map.aspx.

Bracken, B. A., Bai, W., Fithian, E., Lamprecht, M. S., Little, C., & Quek, C. (2003a). *The Test of Critical Thinking*. Williamsburg, VA: Center for Gifted Education, The College of William and Mary.

Bracken, B. A., Bai, W., Fithian, E., Lamprecht, M. S., Little, C., & Quek, C. (2003b). *The Test of Critical Thinking; Examiners Manual*. Williamsburg, VA: Center for Gifted Education, The College of William and Mary.

Bracken, B. A., VanTassel-Baska, J., Brown, E. F., & Feng, A. (2007). Project Athena: A tale of two studies. In J. VanTassel-Baska & T. Stambaugh (Eds.), *Overlooked gems: A national perspective on promising students of poverty* (pp. 63-67). Washington, D.C.: National Association for Gifted Children.

Deno, S. L., Marston, D., & Mirkin, P. (1982). Valid measurement procedures for continuous evaluation of written expression. *Exceptional Children*, 48, 368-371.

Elliot, P. (2014). *Common Core State Standards Focus On Critical Thinking Amid Political Debate*. Retrieved from http://www.huffingtonpost.com/2013/12/01/common-core_n_4367550.html

Hosp, M., Hosp, J., & Howell, K. (2007). *The ABCs of CBM: A practical guide to curriculum based measurement*. New York: Guildford Press.

McKown, C., & Weinstein, R. S. (2008). Teacher expectations, classroom context, and the achievement gap. *Journal of School Psychology,* 46, 235-261.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards*. Washington, DC: Authors.

Paul, R. (1992). *Critical thinking: What every person needs to survive in a rapidly changing world.* Rohner Park, CA: Foundation for Critical Thinking.

Rubie-Davies, C. M. (2007). Classroom interactions: Exploring the practices of high- and low-expectation teachers. *British Journal of Educational Psychology,* 77(2), 289-306.

Shinn, M., (1989). *Curriculum-based measurement: Assessing special children*. Guilford Press, New York.

Shinn, M., Deno, S., Espin. C. (2000). Technical Adequacy of the Maze Task for Curriculum-Based Measurement of Reading Growth. *The Journal of Special Education*, 34(3), 164-172.

Swanson, J. D. (2006). Breaking through Assumptions about Low-Income, Minority Gifted Students. *Gifted Child Quarterly,* 50(1), 11-25.

VanTassel-Baska, J. (1986). Effective curriculum and instructional models for talented students. *Gifted Child Quarterly, 30*, 164-169.

VanTassel-Baska, J. (2003). Content-based curriculum for high-ability learners: An introduction. In J. VanTassel-Baska & C. A. Little (Eds.), *Content-based curriculum for high-ability learners* (pp. 1-23). Waco, TX: Prufrock Press.
VanTassel-Baska, J. (2003), Analyzing differentiation in the classroom: Using the COS-R. *Gifted Child Quarterly*, **35**(1), 43-48.
VanTassel-Baska, J., Avery, L., Struck, J., Fend, A. X., Bracken, B., Drummond, D., Stambaugh, T. (2003). *Classroom Observation Scales-Revised*. Williamsburg, VA. Center for Gifted Education, The College of William and Mary.
VanTassel-Baska, J., Feng, A. X., Brown, E., Bracken, B., Stambaugh, T., French, H., McGowan, S., Worley, B., Quek, C., & Bai, W. (2008). A study of differentiated instructional change over 3 years. *Gifted Child Quarterly,* **52**(4), 297-312.
VanTassel-Baska, J., Johnson, D. T., Hughes, C. E., & Boyce, L. N. (1996). A study of language arts curriculum effectiveness with gifted learners. J*ournal for the Education of the Gifted,* 19, 461-480.
VanTassel-Baska, J., & Stambaugh, T. (2006). Project Athena: A pathway to advanced literacy development for children of poverty. *Gifted Child Today,* 29(2), 58-65.
VanTassel-Baska, J., Zuo, L., Avery, L., & Little, C. (2002). A curriculum study of gifted student learning in the language arts. *Gifted Child Quarterly,* 46, 30-43.
J. VanTassel-Baska, C. Quek, & A. Xuemei Feng (2007). The development and use of a structured teacher observation scale to assess differentiated best practice. *Roeper Review*, winter 2007. Retrieved from http://findarticles.com/p/articles/mi_hb6470/is_2_29 /ai_n29317283/?tag=content;col1

Luke Duesbery was born in Morton on the Marsh, U.K. (1968), obtained a B.A. (Honours) in economics and history from Queen's University, Kingston, Ontario, Canada, an M.Sc. in education from Old Dominion University in Norfolk, Virginia, U.S.A., and a Ph.D. in learning assessment from University of Oregon, Eugene, Oregon, U.S.A. He is currently Associate Professor of Teacher Education and Director for the Centre for Teaching Critical Thinking and Creativity at San Diego State University.

Paul Justice was born in Philadelphia, PA (1966), obtained a B.A. in sociology and geography from Bucknell University, Lewisburg, PA and an MA in linguistics from San Diego State University, San Diego, CA. He is currently a lecturer in the Linguistics Department and an academic adviser in the Division of Undergraduate Studies at San Diego State.

Table 1. Fidelity of Implementation: Presence and Effectiveness of Instruction

Teacher Behaviour [a]	Intervention (*n*=3)	Comparison (*n*=6)	Difference
literary analysis	1.8	0.5	+1.3
persuasive writing	0.0	0.0	+0.0
grammar	0.0	0.3	-0.3
structured questions	1.8	1.3	+0.5
oral communication	2.5	1.0	+1.5
reasoning	0.0	0.5	-0.5
research	0.3	0.0	+0.3
concept maps	0.0	0.0	+0.0
emphasized change	0.8	0.5	+0.3
generalized	0.5	0.0	+0.5
emphasized concepts	1.2	0.2	+1.0

Note a. Scale: 3 = "effective", 2 = "somewhat effective", 1 = "ineffective", = "behaviour did not take place"

Table 2. Percent classroom time behaviour observed in intervention and comparison groups (five pt. scale)

	Intervention	Comparison	Difference
	Mean (*SD*)	Mean (*SD*)	Mean (*SD*)
General Classroom behaviours	**3.5 (1.8)**	**2.4 (1.8)**	**+1.1 (0.0)**
Appeared Challenged	3.3 (1.5)	2.7 (2.0)	+0.6 (-0.5)
Applied learning	5.0 (0.0)	3.7 (1.6)	+1.3 (-1.6)
Appeared thoughtful	4.3 (1.2)	3.0 (1.1)	+1.3 (+0.1)
Evaluated evidence	4.0 (1.7)	2.0 (1.7)	+2.0 (+0.0)
Reflected on learning	0.7 (0.6)	0.7 (1.2)	+0.0 (-0.6)
Diverse self-paced activities	**1.7 (2.5)**	**1.4 (2.3)**	**+0.3 (+0.2)**
Engaged in projects	5.0 (0.0)	1.7 (2.6)	+3.3 (-2.6)
Exposed to tiering	0.0 (0.0)	1.7 (2.6)	-1.7 (-2.6)
Exercised choice	0.0 (0.0)	0.8 (2.0)	-0.8 (-2.0)
Critical thinking	**3.2 (1.9)**	**0.9 (1.8)**	**+2.3 (+0.1)**
Considered purpose	4.3 (1.2)	1.3 (2.2)	+3.0 (-1.0)
Discriminated relevant points	4.3 (1.2)	0.5 (0.8)	+3.8 (+0.4)
Made judgments	2.7 (2.5)	0.8 (2.0)	+1.9 (+0.5)
Analysis and synthesis	**2.2 (1.3)**	**0.6 (1.2)**	**+1.6 (+0.1)**
Compared and contrasted	2.0 (1.0)	1.2 (1.9)	+0.8 (-0.9)
Generalized	3.0 (1.7)	0.7 (1.2)	+2.3 (+0.5)
Synthesized	1.0 (1.0)	0.3 (0.8)	+0.7 (+0.2)
Discovered central ideas	2.7 (0.6)	0.3 (0.5)	+2.4 (+0.1)
Research strategies	**1.4 (1.8)**	**0.0 (0.0)**	**+1.4 (+1.8)**
Gathered evidence	2.3 (2.5)	0.0 (0.0)	+2.3 (+2.5)
Manipulated data	1.0 (1.0)	0.0 (0.0)	+1.0 (+1.0)
Made inferences	1.3 (1.5)	0.0 (0.0)	+1.3 (+1.5)
Determined implications	2.3 (2.5)	0.0 (0.0)	+2.3 (+2.5)
Communicated findings	0.0 (0.0)	0.0 (0.0)	+0.0 (+0.0)
Content in depth	**1.5 (1.8)**	**0.2 (0.8)**	**+1.3 (+1.0)**
Used specialized vocabulary	3.0 (1.7)	1.0 (2.0)	+2.0 (-0.3)
Elaborated on content	1.0 (1.0)	0.0 (0.0)	+1.0 (+1.0)
Asked meaningful questions	1.0 (1.0)	0.0 (0.0)	+1.0 (+1.0)
Dealt with ethical issues	1.7 (2.9)	0.0 (0.0)	+1.7 (+2.9)
Generalized to theme	1.7 (2.9)	0.0 (0.0)	+1.7 (+2.9)
Multicultural content	**0.9 (1.4)**	**0.0 (0.0)**	**+0.9 (+1.4)**
Considered cultural perspective	1.0 (1.7)	0.0 (0.0)	+1.0 (+1.7)
Considered diversity issues	0.7 (1.2)	0.0 (0.0)	+0.7 (+1.2)
Considered social change	1.0 (1.7)	0.0 (0.0)	+1.0 (+1.7)

Note: none and N/A collapsed to value of zero

Table 3. Pre- and post-test scores across critical thinking, writing, and reading.

	Intervention (*n*=41)			Comparison (*n*=87)		
	Pre-score mean (*SD*)	Post-score mean (*SD*)	Effect size *d*	Pre-score mean (*SD*)	Post-score mean (*SD*)	Effect size *d*
critical thinking	18.5 (6.1)	19.8 (6.1)	0.21*	19.2 (6.2)	19.9 (6.6)	0.11*
writing	47.2 (15.3)	44.1 (17.0)	-0.19	46.3 (8.0)	47.4 (17.7)	0.08
reading	15.6 (7.3)	21.6 (10.1)	0.68*	15.5 (8.3)	21.9 (8.9)	0.74*

* $p < 0.01$

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar