



Linking Teacher Learning to Study Success:

An Evaluation of the Walden University Richard W. Riley College of Education and Leadership M.S. in Education with a specialization in Elementary Reading and Literacy

FREQUENTLY ASKED QUESTIONS (FAQS)

1. Why is Walden University conducting this study?

Walden University has engaged in a program of research to determine the effectiveness of its Master of Science in Education with a specialization in Elementary Reading and Literacy (PreK–6). It specifically sought to follow up and build on a 2003-2004 study of Walden teacher graduates in the Peninsula and Tacoma, Wash., school districts. Walden sought to identify districts with high concentrations of Walden graduates where it could study how well the students of Walden graduates performed compared to students of a defensible comparison group of teachers.

2. When was the research conducted?

The study was conducted from May 2008 through April 2009, using data from 2005 through 2008.

3. Who conducted the research?

The study was conducted by Arroyo Research Services (ARS), an education-research, measurement and consulting firm. The principal investigators were Kirk Vandersall and David Stuit.

Kirk Vandersall is Managing Director of ARS. Vandersall leads ARS' strategic partnership with TrueNorthLogic, providing assessment development, evaluation-and research-advisory services. His work includes a two-year evaluation of the Texas Dropout Recovery Pilot Program for the Texas Education Agency; a multi-year evaluation of the Columbus Public Schools LACES Reading Program (Ohio); the Collier County Migrant Education Program (Florida); the Electronic Portfolio Connection at five universities in Vermont; the simSchool online teacher-training simulation (CurveShift and the University of North Texas); ETIPS Educational Leadership Cases at the University of Virginia; an Enhancing Education Through Technology project in Dayton, Ohio; an innovative, technology-based early childhood educator-professional-development program under development by the Arizona School Services through Education Technology (ASSET) program;

and online M.S. in Education programs offered through Walden University. Additionally, Vandersall led the development of a Curriculum Management Plan for the Baltimore County Public Schools, where he is developing a Comprehensive Assessment Plan and directing ARS advisory services. He leads the Effective Practices Team of the National Consortium of Interpreter Education Centers at Northeastern University; directs Arroyo Research Services' No Child Left Behind (NCLB) evaluation technical-assistance efforts for members of Florida's East Coast Technical Assistance Center (ECTAC); is the facilitator of the Florida Evaluation Working Group; and recently designed and delivered evaluation services for the NCLB Regional Technical Assistance Centers for Regions III and IV in Florida. Additional recent client work includes student- and teacher-data analysis for SchoolNet, Inc.; review of English Language Development programs for a large southern California school district; consulting services for American Productivity and Quality Center (APQC); and research-design services for Gallup's education division. Vandersall also served as special advisor to the APQC K–12 Professional Development Benchmarking Study.

Vandersall has more than 15 years of experience in leading and writing evaluations and policy studies at the federal, state, and local levels, and providing a broad range of professional services for education organizations. He is a contributing author on state-education policy in *Strategies for Equity: Creating Productive Schools in a Just Society*, edited by Marilyn Gittell, and has published and presented at national conferences and seminars on educational measurement and evaluation, education technology, state education policy-making, state school finance, and educational equity. He specialized in state and local policy studies at the Howard Samuels State Management and Policy Center in New York City, where he led education-policy studies throughout the country. Prior to forming Arroyo Research Services, Vandersall was a founding partner of Metiri Group, a national education-technology consulting firm, where he also was Practice Leader for Metiri Group Research and Evaluation. In that capacity, he managed a three-year evaluation of the Ohio SchoolNet Telecommunity; was engagement manager for university and foundation initiatives in Vermont, Illinois, Minnesota, Pennsylvania, Florida, Virginia, and Iowa; advised the U.S. Department of Education and State Educational Technology Directors Association (SETDA); and provided research-and-evaluation services for state and national clients. Vandersall was a key contributor to the enGauge assessments developed by Metiri Group for the North Central Regional Educational Laboratory, which included the International Society for Technology in Education's National Educational Technology Standards for Teachers (ISTE NETS-T) aligned assessments of educator proficiency. He also performed the technical validation studies for the SETDA-Profile of Education Technology Integration (PETI) surveys for teachers, principals and district administrators, and managed the technical development of Metiri Group's online assessment products, including its version of the SETDA-PETI products, as well as survey administration tools for students and teachers.

Vandersall established an office of assessment and evaluation for a large urban school district in Los Angeles County; evaluated a major K–12 math-and-science initiative funded by the National Science Foundation; taught graduate-level policy analysis and evaluation; and has been a research-and-evaluation consultant for foundation, nonprofit and corporate projects in education, healthcare and urban policy.

David Stuit is an ARS Associate. Stuit is a Ph.D. candidate in the Leadership, Policy, and Organizations department of the Peabody College of Education at Vanderbilt University. Stuit has expertise in statistical analyses and field-research design. At ARS, he assisted in a longitudinal evaluation of Florida migrant students; co-designed with the University of Virginia field studies of the principal training software Educational Theory into Practice; and assisted in evaluating the effectiveness of Texas' statewide dropout recovery program. While at Vanderbilt, Stuit was a fellow in the U.S. Department of Education's Institute for Education Sciences' Experimental Education Research Training program, designed to train the next generation of education researchers in the skills necessary to execute large-scale randomized field trials and analyze longitudinal data sets using advanced statistical methods. Prior to Vanderbilt, Stuit conducted research for the teacher quality division of the Education Commission of the States (ECS). He has served as a member of the student editorial board of the American Educational Research Association's (AERA) *Educational Evaluation and Policy Analysis* (EEPA) journal and has reviewed article submissions for the *Peabody Journal of Education*. He is a member of the AERA; the American Education Finance Association; the Association for Public Policy Analysis and Management; and the Society for Research on Educational Effectiveness. Stuit began his career as a teacher in the Cherry Creek School District in Denver, Colo.

4. Why was Tacoma selected as the study site?

Walden sought to identify school districts that had high concentrations of Walden University M.S. in Education graduates with a specialization in Elementary Reading and Literacy (PreK–6), had reasonably strong assessment programs, and were willing to allow Walden University access to student and teacher data for the purposes of conducting the study. The research team approached several districts in Georgia, New Mexico and Washington. Tacoma best fit this profile and was willing to participate. Moreover, Tacoma was one of two districts that participated in a partnership program with Walden in the 2003–2004 school year that was the focus of the original PenTac Study.

5. Why was DIBELS Oral Reading Fluency (ORF) selected as an outcome measure?

The DIBELS Oral Reading Fluency (ORF) was used as an outcome measure because it was the best available reading measure that the district administered in every elementary grade (grades 1 – 5). One advantage of using the DIBELS ORF is that the district administers the test three times each year: once in the fall, once in the winter, and once in the spring. This allowed the researchers to calculate the growth in oral reading fluency that students make during the school year as the difference between a student's spring test score and fall test score. There is substantial empirical evidence to suggest that oral reading fluency is a reliable measure of students' overall reading competence. Adams (1990) considers oral reading fluency the most salient characteristic of skillful reading. According to Lynn Fuchs and associates (2001):

“...oral reading fluency represents a complicated, multifaceted performance that entails, for example, a reader's perceptual skill at automatically translating letters into coherent sound representations, unitizing those sound components into recognizable wholes and automatically accessing lexical representations, processing meaningful connections within and between sentences, relating text meaning to prior information, and making inferences to supply missing information. That is, as an individual translates text into spoken language, he or she quickly coordinates these skills in an obligatory and seemingly effortless manner, and because oral reading fluency reflects this complex orchestration, it can be used in an elegant and reliable way to characterize reading expertise.” (p. 240)

Numerous studies have shown oral reading fluency is highly correlated with other measures of reading comprehension (see, for example, Fuchs et al., 2001; Daane et al., 2005). Fuchs et al. (1988) found that oral reading fluency is more highly correlated with performance on standardized tests of reading comprehension than more direct measures of reading comprehension.

The specific DIBELS Oral Reading Fluency measure has also been shown to have a moderate to strong association with state standardized reading assessments (Barger, 2003; Buck & Torgenson, 2003; Good et al., 2001; Shaw & Shaw, 2002, Vander Meer et al., 2005; Wilson, 2005). Riedel and Samuels (2007) found that first grade students' DIBELS ORF scores were good predictors of their reading comprehension at the end of first grade and second grade. A recent study by Roehrig et al. (2008) found moderate correlations between third grade students' ORF scores and reading comprehension scores on the FCAT-SS and SAT-10.

Adams, M. J. (1990). *Beginning to Read: Thinking and Learning about Print*. Cambridge, MA: MIT Press.

Fuchs, L.S., Fuchs, D., Hosp, M.K. & Jenkins, J.R. (2001) Oral Reading Fluency as an Indicator of Reading Competence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, 5(3), 239-256.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The Validity of Informal Measures of Reading Comprehension. *Remedial and Special Education*, 9(2), 20–28.

Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina End of Grade reading assessment (Technical Report)*. Asheville: North Carolina Teacher Academy.

Buck, J. & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test (FCRR Technical Report No. 1)*. Tallahassee: Florida Center for Reading Research.

Good, R.H. & Kaminski, R.A. (2002). *Dynamic Indicators of Basic Early Literacy Skills (6th ed.)*. Eugene, OR: Institute for the Development of Educational Achievement.

Shaw, R. & Shaw, D. (2002). *DIBELS Oral Reading Fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP) (Technical Report)*. Eugene, OR: University of Oregon.

Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instruments to Measure Standards (AIMS)*. (Technical Report). Tempe, AZ: Tempe School District No. 3.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading (Technical Report)*. Eugene, OR: University of Oregon.

6. Why wasn't the Degrees of Reading Power assessment used in the follow-up study?

The Degrees of Reading Power assessment was not consistently administered across all students in all grade levels of interest during the years of the study. Tacoma Public Schools changed the target population for DRP, therefore making DIBELS the most appropriate assessment with the most observations across the years of the study, resulting in maximum use of all program graduates and comparison teachers.

7. How were teachers and students selected for the study?

The teachers included in the initial follow-up study were those who participated in the original PenTac study and were still working in the Tacoma school district. This included nine Walden M.S. in Education graduates with a specialization in Elementary Reading and Literacy (PreK–6) and six comparison group teachers. To select teachers who graduated from the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) for the expanded follow-up study, the research team identified all teachers who worked in grades 1 through 5 in the Tacoma school district from 2006 to 2008 and who graduated from the Walden M.S. in Education program before the fall of 2007. Each Walden teacher was matched to a non-Walden teacher who taught the same grade, held a master’s degree from a different institution, had a similar level of teaching experience, and in most cases worked in the same school. District records were then used to link the teachers to their students in each of the three years of data.

8. How comparable were the treatment and control groups?

The teacher and student characteristics of the expanded study sample are shown in the following tables drawn from the White Paper:

Table 6: Teachers in the Expanded Tacoma Study

	Walden	Matched
Pct. Female	83%	83%
Pct. Master’s	100%	100%
Pct. 1–2 Years’ Experience	0%	0%
Pct. 3–5 Years’ Experience	8%	6%
Pct. 6–10 Years’ Experience	26%	11%
Pct. 11–20 Years’ Experience	0%	0%
Pct. Over 20 Years’ Experience	66%	82%
Teachers	35	35

Table 7: Students in the Expanded Tacoma Study

	Walden	Matched
Asian	11%	11%
Black	24%	19%
Hispanic	13%	11%
White	35%	43%
ESL	7%	4%
Special Ed.	9%	10%
Student Obs.	1834	1994

9. Did teachers and students know they were being studied?

No. While the original PenTac participants all knew they were in a study, and received specific incentives for participating, this follow-up study was an observational study only. That is, the research team used existing district assessment and teacher data to construct the dataset used for analysis. This low-impact approach was important to the district, which did not approve more intrusive classroom observations, teacher surveys, or other data gathering.

10. How did the research team obtain data for the study?

Data for the study were obtained from three main sources. 1) Most data was provided by Tacoma Public Schools, from their district data warehouse, under a formal Agreement for Conducting District-Approved Research. This included student demographics, teacher demographics, and student assessment performance on multiple measures from 2004 to 2008. Some teacher experience and education information was also obtained from TPS. 2) Walden University provided information in their graduates. 3) MDR provided information on degrees earned for all TPS teachers.

11. Did the control group include graduates of other online programs?

The name of the university and major field of study of the non-Walden graduates are unknown. ARS secured data on the districts' teachers from MDR, this data was limited in that it only included a Yes/No indicator of whether the teacher held a master's degree or not.

12. The treatment group includes graduates from programs other than literacy programs. How did this affect the research outcomes?

Literacy instruction is a key responsibility of primary and elementary school teachers, and the extent to which teachers succeed in teaching literacy is the outcome studied in this paper. The paper does not seek to compare Walden's M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) to other elementary literacy programs; it seeks to compare the Walden program to other master's degree programs. Indeed, the emphasis on literacy and reading, plus the balance of the Walden experience, are the intervention.

Because of this, the teacher's master's degree is not an attribute of the teacher that we want to make exactly comparable for the treatment group (Walden teachers) and control group (non-Walden teachers). Rather, it is the "treatment", or intervention that we are evaluating to determine if it associates with increased teacher effectiveness, as measured by student reading achievement. To evaluate the effectiveness of the Walden M.S. in Education specialization in Elementary Reading and Literacy (PreK–6), we needed to specify a control treatment with which to compare it. The control treatment one chooses is driven by the question one wants to answer and/or is able to answer. Among other questions, researchers may want to know: (1) Does the treatment perform better than the average of all the other treatments used in the field? (2) Does the treatment perform better than those treatments that are most commonly used? (3) Does the treatment perform better than other similar treatments?

This study sought to find evidence to answer the first question: Does the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) program make teachers more effective at reading instruction than the average of other master’s degree programs? This is the question that was possible to answer given that we did not have data on the name of teacher’s graduate institutions or information on their major field of study. Moreover, it is an appropriate question to answer before proceeding to determine if the Walden program is more effective than other elementary and literacy master’s programs and/or more effective than other distance-learning programs. A common malpractice within education research is for researchers to jump to answering the third question before answering the first two questions (this is currently the topic of discussion for reforming the federal education research agenda). The goal of early-stage research, such as this, should be to compare the treatment to very different control conditions to see if you can identify a “main effect” of the program. Once a main effect has been established, then it makes sense to proceed to comparing the Walden program to similar other programs—an effort that will likely require larger sample sizes in order to identify statistically significant differences.

What threats to the validity of the findings stem from comparing the Walden teachers to teachers with master’s degrees in other areas? It is important to note: it is not a problem that treatment teachers received a degree specializing in elementary reading and literacy and control teachers did not. However, it may be a problem if those teachers that sought out degrees with specializations in elementary reading and literacy were also those teachers who were naturally more interested in reading instruction and motivated to improve their reading instruction. If this is the case, these teachers may be more effective reading instructors regardless of whether or not they attend the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) program. This is one of many threats to the validity of the results that stem from self-selection into treatment groups. These threats can only be overcome via an experiment and this is why we make no argument with these results that there is a causal relationship between the Walden program and instructional effectiveness.

13. How many data points or observations were analyzed for this research?

The two main analyses described in the report are referred to as the Initial Follow-Up Study and the Expanded Follow-Up Study. In the Initial Follow-up study, we compare the performance 436 students of 9 Walden teachers to the performance of 351 students of 6 comparison group teachers over three years (2006, 2007, 2008). In the Expanded Follow-up Study, we compare the performance 1,865 students (grades 1 -5) of 35 Walden ERL graduates to the performance of 1994 students of 35 control group teachers over three years (2006, 2007, 2008).

	Treatment Group Students (Teachers)	Comparison Group Students (Teachers)	Total Observations
Initial Follow-Up Study	436 (9)	351 (6)	787 (15)
Expanded Follow-up Study	1,865 (35)	1,994 (35)	3,859 (70)

14. Are the sample sizes within normal range for this type of research?

The ideal sample size for a given study depends on a number of factors, including the type of intervention that is being studied, the unit of analysis (e.g., students, teachers, schools). The size of the teacher sample is most important in this analysis because the treatment we are evaluating is a “teacher-level” treatment. The teacher sample size used in the initial follow-up study (15) is small by conventional standards. A general guideline that many researchers follow for using HLM on data at two levels (in this case students and teachers) is that there should be at least 30 teachers in a sample, although this rule is fairly arbitrary and often violated. Our expanded study satisfies this rule because it uses 75 teachers and is a suitable sample size for the research questions being addressed.

15. How did the research model take into account district reading programs, e.g., Reading First, Title I, or other programs and characteristics that might vary by school?

The regression model included an indicator of whether the school received Reading First funding to support reading professional development and classroom-based reading assessments. In addition, the model included an indicator of whether the school had a full-time reading instruction facilitator, which are most likely also funded by Title I grants. All of the schools designated as school-wide Title I had an instructional facilitator and none of the schools that were not designated as school-wide Title I had an instructional facilitator. Therefore, the indicator of instructional facilitator also captures any additional effects of additional Title I support on student reading achievement.

Controlling for school effects was also an important consideration in the study. Two strategies were used to ensure (as best possible) that the difference between the Walden teachers and the non-Walden teachers was not due to differences in the instructional conditions, programs, principals, or other factors that might influence the reading achievement of all students in the school. First, we attempted to match each Walden teacher to a teacher in the same school who held a master’s degree. This strategy is an attempt to balance the school effects in the sample. We were successful in finding matches for 16 of the 35 Walden teachers. For the remaining 19 teachers, we needed to look for matches outside the school, because there was not a teacher with a master’s degree who taught the same grade in their school. Since we were not able to perfectly balance the schools in the sample, we used a second strategy. Our second strategy was to control for other factors of the school’s reading program in the regression model. We included an indicator if the school had a full-time reading instructional facilitator, because these schools may experience a raise in performance because of this additional resource. Similarly, we included an indicator of the school received Reading First funds, which go towards professional development and reading assessment tools in the schools and may also lead to an increase in performance.

16. How do we explain the differences in effectiveness of the M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) in teachers of earlier elementary grades (K–3) as compared to the higher grades (4–5)?

The results of the study did not show that there was a systematic difference in the effectiveness of the M.S. in Education with a specialization in Elementary Reading and Literacy (PreK–6) for early elementary grades (K–3) and higher grades (4–5). The results in Table 8 show that the effect of the program was different at each grade level, with positive effects found in grades 1, 2, 3, and 5, and a negative effect found in grade 4. The estimated differences in the number of words per minute between Walden teachers and non-Walden teachers were largest in grades 1 and 2. However, when the differences are represented as percentages of the average control group’s fall to spring ORF gain, the magnitude of the fifth-grade effect is actually largest.

The reason why the fourth-grade effect is negative is unknown and left to our speculation. It may be that some or all of the fourth-grade teachers did not respond to the program in the same way as the teachers in other grades. It may be that the specific instructional strategies needed in fourth-grade reading instruction are not sufficiently addressed with the M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) curriculum.

As with the observed positive effects of the Walden program, the negative fourth-grade effect may be due to limitations of the research design, namely unobserved differences in the students, teachers, and/or schools and comparison samples that could only be addressed with randomization. For example, maybe the comparison group of fourth-grade teachers was comprised of teachers that had gone through intensive professional development in literacy instruction (a factor we did not know), or maybe the comparison group contained a high concentration of naturally gifted reading teachers.

17. How does this research apply to recently enrolled students, who are enrolled in the revised program? Are there any recent changes to the M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) that would affect these results in future graduates? If yes, what are they?

All teachers in this study graduated from the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) prior to the launch of the revised program. Therefore, the study was an evaluation of the previous program. The extent to which findings of the evaluation generalize to the effectiveness of the revised program depends on the extent of the changes that were made to the program. Our understanding is that the revised program maintained most of the core features of the original program, in which case one would expect the effect of the original program to generalize to the revised program.

18. The Expanded Follow-Up Study seems to show smaller gains than the original PenTac study or the Initial Follow-up Study. Why? And how should this be interpreted?

The findings of the original and follow-up PenTac study provided suggestive evidence that the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) associated with positive gains in student reading achievement. However, both studies had substantial limitations in their design, which called the results into question. The findings of the original study were questionable because of the methods used to compare the performance of Walden teachers to comparison group teachers were incorrect. The findings of the PenTac follow-up study were subject to question because the sample only included 15 teachers, which was too small to allow us to use the appropriate statistical techniques to control for some of the differences in the student, school, and teacher characteristics between the nine Walden teachers and the six comparison group teachers.

Conducting the expanded study with a sample of 35 Walden teachers and 35 comparison group teachers provided a larger sample that allowed us to conduct a more rigorous analysis that controlled for differences in the characteristics of the teachers, the characteristics of the schools, and the characteristics of the students. Consequently, the findings of the expanded study are more reliable. That is, they provide stronger suggestive evidence to support the argument that the Walden program works. It is a positive sign that the findings of the expanded study align with those of the two earlier studies.

19. How did the authors calculate the data in Table 8: Predicted Differences in Fall to Spring Oral Reading Fluency Gains?

The data in Columns 2–4 of Table 8 come from the results of the regression analysis. The data in Column 5 (the difference as percent of average control group fall to spring gains) is found by dividing Column 4 by Column 3. The data in Column 6 (additional weeks of instruction) is found by dividing the average fall to spring gain of the control group (Column 3) by 35 (the number of weeks in a typical school year) to yield the average weekly gain in oral reading fluency of the control group during the school. The “Walden effect” in Column 4 is then divided by the average weekly gain to determine the number of additional weeks a control group student would need to make up the difference.

Table 8: Predicted Differences in Fall to Spring Oral Reading Fluency Gains

	Average Fall to Spring Gain, Walden	Average Fall to Spring Gain, Control Group	Difference in Fall to Spring Gains	Difference as % of Average Control Group Fall to Spring Gains	Additional Weeks of Instruction Needed to Make Up Gain (35 Weeks)
Grade 1	67.2	61.8	5.4	9%	3.1
Grade 2	51.1	47.5	3.6	7.5%	2.7
Grade 3	37.7	36.2	1.5	4%	1.5
Grade 4	35.2	36.3	-1.1	-3%	-1.1
Grade 5	19.9	17.7	2.2	12%	4.4
Weighted Average	41.2	38.3	2.95	8%	2.7

- Difference in Fall to Spring Gains = [Avg. Fall to Spring Gain Walden]–[Avg. Fall to Spring Gain Control]
- Difference as percent of Avg. Control Group Fall to Spring Gains = [Diff. in Fall to Spring Gains]/[Avg. Fall to Spring Gain Control Group].
- Additional Weeks of Instruction Needed = [Difference in Fall to Spring Gains]/([Avg. Fall to Spring Gain Control Group]/[Presumed Weeks of Instruction (35)])

20. How should the regression results in the Appendices be read and interpreted?

The tables in the appendix are provided as reference for those with a technical interest in the specification and results of the regression model. The table below shows the results of the expanded analysis as a reference. The outcome being predicted was the fall to spring gain on the DIBELS ORF. Each factor in the leftmost column of the table was a term used in the analysis because we expected it to associate with ORF gains. The numbers in the columns to the right of each factor represent the amount of additional gain attributable to that factor when all other factors are held constant.

The number we are most interested in is that next to the first term, Walden, which is an indicator of whether the teacher was a Walden graduate. This value tells us the average effect of the Walden teacher on students' ORF gains when all other factors (race, ESL, special education, Reading First, Instructional Facilitator, etc.) are held constant. It is our "treatment effect". The second column presents the results of a regression model where we assume the "Walden effect" on ORF gains is constant across all grades. That is, we assume a first-grade Walden teacher has the same effect on ORF gains as a fourth-grade Walden teacher, etc. This model is used to give us an "average" effect of the Walden teachers over all students in all grades in the sample, which we found to be 2.95 additional words per minute. The third column presents the results of a regression model where we test to see if the "Walden effect" is different across grades. That is, we test to see if the effect of first-grade Walden teachers is statistically different than the effect of second-, third-, fourth-, or fifth-grade Walden teachers. In this table, the number for Walden (5.41) is actually the "Walden effect" for first-grade teachers. To get the "Walden effect" for the other grades, one needs to add 5.41 to the numbers next to the terms: second grade* Walden, third grade*Walden, fourth grade*Walden, fifth grade*Walden. For example, to calculate the Walden effect for fifth-grade students, one would add -3.25 to 5.41 and arrive at 2.16.

Grade-level indicators are included to control for the normal level of gains made within each grade level, which differs by grade level for the ORF gains. The numbers presented for second grade, third grade, fourth grade, and fifth grade are negative because they are in reference to the average first-grade gain, which is captured in the constant term. The constant is the average value in the sample after all other factors are controlled. In this case, it is interpreted as the average gain for first-grade students in the comparison group after all other factors are held constant. For example, the constant value of 61.88 shown in Column 3 is the average ORF gain for first-grade students of comparison group teachers, while controlling for student factors (ethnicity, ESL status) and school factors (Reading first, instructional facilitator). The model also includes year indicators to account for differences in overall performance each year.

The asterisks at the bottom of the table indicate the p-value or level of statistical significant for each coefficient. One asterisk indicates that the estimated difference in ORF gains for a given factor has less than a 10 percent probability of showing up in the data by chance. Two asterisks indicate the difference has less than a 5 percent probability of showing up in the data by chance. Three asterisks indicate the difference has less than a 1 percent probability of showing up in the data by chance.

Table A-2: Expanded Follow-up Study Regression Results

	ORF Gains			
	Constant Effect Model		Grade-Varying Effect Model	
Walden	2.95	**	5.41	**
Black	-2.69	**	-2.66	**
Hispanic	-2.69	**	-2.66	*
White	-1.33		-1.46	
ESL	12.16	**	-12.18	**
Special Education	-12.58	***	-12.57	**
Reading First	5.17	**	5.24	***
Instructional Facilitator	-4.44	**	-4.39	**
2nd Grade	-15.134		-14.35	***
3rd Grade	-27.32	***	-25.61	***
4th Grade	-28.70	***	-25.66	***
5th Grade	-45.57	***	-44.07	***
2nd Grade* Walden			-1.8	*
3rd Grade* Walden			-3.88	*
4th Grade* Walden			-6.39	*
5th Grade* Walden			-3.25	*
2007	1.80		-1.80	
2008	0.77		-0.78	
Constant	62.99	***	61.8	***
No. of Teachers	70		70	
No. of Student Observations	3,859		3,859	
Chi-Square Test of HLM vs. OLS	27.75	***	23.75	***

* $p \leq .10$. ** $p \leq .05$. *** $p \leq .01$.

21. What is Walden Education Outcomes Working Group? Did they have any influence on the outcomes of the study?

The Walden Education Outcomes Working Group is an internal Walden University work team comprised of senior administrators committed to supporting ongoing research regarding Walden University and its programs. The group developed a research program and set of guiding questions, marshaled budgetary resources, and commissioned this study. The Outcomes Working Group had no involvement in the conduct of the study or the development of findings. The Outcomes Working Group reviewed near-final versions of the White Paper, providing comments, feedback, and requests for clarification of terms and audience, after the analysis was complete.

22. To what do you attribute the fourth-grade results? Is that an indication of a weakness in the Walden program?

The reason why the fourth-grade effect is negative is unknown and left to our speculation. This study was conducted without the benefit teacher implementation measures, measures of teacher approaches to instruction, or measures related to the role of teacher education in the preparation of lessons. It may be that some or all of the fourth-grade teachers did not respond to the program in the same way as the teachers in other grades. It may be that the specific instructional strategies needed in fourth-grade reading instruction are not sufficiently addressed with the elementary reading and literacy curriculum. Moreover, as with the observed positive effects of the Walden program, the negative fourth-grade effect may be due to limitations of the research design, namely unobserved differences in the students, teachers, and/or schools and comparison samples that could only be addressed with randomization. For example, the comparison group of fourth-grade teachers may be comprised of teachers that had gone through intensive professional development in literacy instruction (a factor we did not know), or the comparison group may simply contain a high concentration of naturally gifted reading teachers.

23. Is there a fall-to-spring WASL test that would be in same time frame as the study, or only spring to spring? Does that disparity in time measured have an effect on the results?

The WASL is a test of the state's learning standards and is administered annually in the spring to all public school students in Grades 3 through 8. Because WASL testing does not begin until third grade and the majority of the identified Walden teachers are working in Grades 1 and 2, this severely limited our ability to use WASL as the major outcome of the study. Additionally, because we needed to calculate a spring-to-spring gain score for each student, and since second-grade scores were not available to calculate gain scores for third-grade students, the WASL sample was limited to the 12 Walden teachers working in Grades 4 and 5. These factors led to the limited utility of the WASL for the purposes of the study discussed in the White Paper.

24. The findings of this study do not allow for a claim of causation between the completion of a Walden degree and teaching effectiveness. Why?

A claim of causation would only be possible if we could rule out all the alternative explanations for why Walden elementary reading and literacy teachers were shown to have larger average ORF gains than the comparison group teachers. There are a number of alternative explanations that we were unable to rule out with this study. Many of these alternative explanations stem from the fact that the Walden teachers self-selected treatment (i.e., the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6)). It is possible that the teachers who seek out a Walden M.S. in Education degree with a specialization in Elementary Reading and Literacy (PreK–6) have certain qualities or skills that would make them more effective teachers than the comparison group teachers even if they did not participate in the program. For example, if those who seek out the Walden M.S. in Education degree with a specialization in Elementary Reading and Literacy (PreK–6) have, on average, more motivation to improve student literacy than those in the comparison group, this research design would incorrectly attribute the effect of this extra motivation on student ORF gains to the Walden program. This threat to the validity of the findings is called "selection bias." The only way to overcome selection bias is to conduct an experiment where teachers are randomly assigned to receive the "Walden treatment" or receive a control treatment. Given that our study was a secondary data analysis and could not employ a randomized design, we leveraged the available data as best possible to find the association of the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6) to teacher effectiveness and provide suggestive evidence on the effect of the Walden M.S. in Education program with a specialization in Elementary Reading and Literacy (PreK–6).

Media Inquiries

Walden University has several spokespersons available to speak with the media about this study, the Richard W. Riley College of Education and Leadership and its programs, including reading and literacy, and other related topics.

For more information, or to schedule an interview with a Walden expert, please contact:

Jerry Sweitzer

Public Relations Manager

Walden University

410-843-6576

jerry.sweitzer@waldenu.edu