

Looking Inside the Education Black Box

William L. Johnson, Ed.D.; Annabel M. Johnson, Ph.D.; and Jared W. Johnson, B.S.

This document was from a presentation at the Annual Meeting of the Science Teachers Association of Texas (STAT), Dallas, Texas, November 21-23, 2019.

Seven thousand students a day are dropping out of the public schools in the United States at a rate of one student every 26 seconds, but the presentation was not only about students dropping out of the public schools. It was what could be done to help alleviate student dropouts and increase student achievement. The major topics covered in the presentation follow: A Brief History of Student Academic Issues; Robert Marzano's 2003 Research Findings; John Hattie's Visible Learning (2009) and Visible Learning for Teachers (2012); and High-Stakes Testing.

The presenters looked at student achievement and accountability, the major reasons students failed, and the predictors of student dropouts (including the EO919 Texas AT-RISK-INDICATOR-CODE reference). The Essentials of Cross-Battery Assessment (Flanagan & Ortiz, 2001) was also reviewed since children and youth with disabilities are a major part of public-school education.

An article in the September 28, 2018 issue of the Houston Chronicle (Texas) noted that a majority of students at the top-rated schools in Texas were likely to need remedial course work after entering college. "More than 900 high schools in the state received the equivalent of an A or B rating from the state in 2018. But in two-thirds of those schools, the majority of students were failing to score high enough on the SAT or ACT to be considered ready for college. This indicated those students would need remedial work in college" (Matos, 2018, p.1).

Having taught special education for several years and been a diagnostician, the lead presenter's special education students passed Texas' TAKS and STAAR tests at levels normally expected of regular students. This often occurs with teachers who hold special education credentials and know how to teach these students. In the 2019-2020 school year, the presenters used an ANOVA test to compare the achievement levels of various groups of "special students" in their classes: Limited English Proficient (n=12); Inclusion (n=20); Bilingual (n=6); English as a Second Language (n=6); Gifted and Talented (n=10); Dyslexic (n=21); and 504 (n=21). The p value was calculated under the assumption that there was no true effect or no true difference. The small p value provided evidence that the research findings were true.

We ran the ANOVA analysis with a posthoc Bonferroni correction to control the false discovery rate. We accepted the null hypothesis regarding the significance of difference between the seven means for the seven groups: ($F(6,89) = 0.262$ $p < 0.05$). We also examined the range of plausible values for the parameter (population mean) and were confident that the population mean lay within the upper-and-lower bounds of the estimated interval (Ellis, 2016; Thompson, 2002).

In the last part of the introduction, the presenters discussed the use of a paced-curriculum calendar. The calendar listed topics and student expectations (SEs) to be covered each six weeks as well as the number of days to cover the material. With the pacing document and standards-based assessments, we retested each test and entered 3000-4500 student grades each six weeks of the academic year. We estimated standards based assessment increased student achievement about 20%. This was a significant factor in our student's achievement. The presenters next discussed the research literature pertaining to student achievement.

Robert Marzano's Research

Following the basic Athenian ideas about education, not much changed for the next 2000 years (Thut, 1957). But in the last few decades, researchers have begun conducting real, evidence-based research into what really worked and what really didn't work in school, teaching and learning to increase student learning (Johnson & Johnson, 2016, 2018).

The author of more than 30 books and 150 articles, Robert Marzano has been the pioneer of large research-based studies. His high reliability school framework has shown how best practices work together to impact student achievement: a safe and collaborative culture, effective teaching in every classroom, a guaranteed and viable curriculum, standards-referenced reporting, and competency-based education (Marzano, Warrick & Simms, 2014). In his 2003 book, "Classroom Management that Works," Marzano found the teacher was the most important factor affecting student learning, and student classroom management most affected student achievement (Marzano, Marzano & Pickering, 2003).

Following Marzano's research, from several years of teaching, we have observed that schools are not buildings, time tables and technology. At the most fundamental level, schools are about relationships. In their study of school districts involved in substantive change, Spillane and Thompson (1997) referenced the research of economist, J. Coleman (1998), that "local capacity" for substantive change was based on three things: physical capacity (financial resources), human capital (administrators and teachers) and social capital (internal and external district relationships).

Ball and Cohen (1995) wrote that physical capital was observable, but human capital, represented in the skills and knowledge of individuals, was even less tangible. Finally, social

capital, represented in the relationships among persons, is even less tangible. Ball and Cohen noted that human and social capital were essential elements in understanding what made a school exemplary.

Writing in the September 2004 issue of *The School Administrator*, Jim Peters, former superintendent of the Shelby, Michigan, public schools asked which elements made the difference in a highly successful and a less successful school: involved parents, socioeconomic status (SES) of students, a highly trained staff or a caring environment where everyone felt connected and respected. Peters noted the answer was all of the above; however, he commented that parental involvement and SES were outside the schools' control. He explained how his district received numerous state awards for exemplary student achievement. He noted that his district's highest achieving schools were those where the students and teachers trusted, respected and cared about each other. This also was true for his principals and their staffs: they trusted, respected and cared for each other and worked toward the same goals focused on student success.

As Peters' district made connections and relationships a district-wide focus, more of its schools received Michigan's highest awards for student achievement and improvement. Overall, the administrators worked with their staffs to support three district-wide objectives: every student was greeted upon entering the classroom; every teacher posted a daily agenda in the classroom; and every teacher made at least two positive calls or parent contacts each week. His school administrators also made greeting students a priority, had agendas for all meetings and made positive contacts with students, staff and parents.

The lessons learned in this Michigan district make a lot of sense to us. We have observed that successful teachers get to know their students, build positive relationships and work to make their classes enjoyable places to work and learn. Their students have excelled both academically and socially.

Having observed the results of positive school environments, Johnson and Johnson (2000) studied the second-order factor structure of the Charles F. Kettering School Climate Profile (CFK), a popular measure of school climate. The authors found that the higher-order structure was composed of two factors: cognitive and affective components. The emotional components were comprised of respect, trust, morale and student input questions. These emotional components were essentially the characteristics that Peters observed in his district's exemplary schools. These characteristics are fundamental relationship components that foster a sense of school community, cooperation and student achievement. We greeted students at the door, put daily agendas on the board and generally made about 75 parent contacts each six weeks. We found relationship management was a very important dimension in student achievement.

John Hattie's Research

John Hattie, professor and director of the Melbourne Education Research Institute at the University of Melbourne, Australia, has been praised for ushering in a new era of school reform and bringing education research to classroom teachers. His 1992 pioneering synthesis of 134 meta-analyses demonstrated the practical utility of calculating average effect sizes across school factors like methods of instruction and learning strategies. Based on his 15 years' research synthesis of more than 800 meta-analyses of 50,000 research articles, 150,000 effect sizes and 240 million students primarily in North America, Hattie's first book, "Visible Learning" (2009),

represented evidence-based research into what actually worked in schools to improve student learning.

His second book, “Visible Learning for Teachers” (2012), took the next step by explaining how to apply the principles of “Visible Learning” to classrooms. His book had user-friendly summaries of the most successful educational interventions and instructional strategies impacting student learning. Hattie found six major influences on student achievement and the contribution of each to student learning (Hattie, 2012, p.14). That summary follows below:

Student +15%	(ES +0.39, +1.0 year)
Home +12%	(ES +0.31, +0.78 year)
School +9%	(ES +0.23, +0.6 year)
Teacher +18%	(ES +0.47, +1.2 year)
Curricula +17%	(ES +0.45, +1.1 year)
Teaching +17%	(ES +0.43, +1.1 year)

The key to Hattie’s research has been to study the influences and apply those needed for one’s classes. For example, if a teacher’s class scored at the seventh-grade level academically, the learning process would equate to Piaget’s third level of concrete operational capability and strategies catalyst (Hattie, 2012 p. 105). That level of students (not necessarily age) would need instruction in the use of strategies. Interestingly, less than 50% of American 11th -and- 12th grade students achieve at Piaget’s top (fourth) level of formal operational capability. Again, the most important factor influencing learning (not necessarily teaching since that’s another different topic) is what the student already knows).

Teachers should study the achievement literature and apply the needed influences in their classes. This is the challenge: applying the research in the classroom. Perhaps the greatest challenge today in education is closing the achievement gap. We know the differential spread in school is about one year for each grade; thus, by the 6th grade, there is a six-year academic differential. By the 10th grade, there is a 10-year academic differential. Therefore, what would be an excellent teaching strategy or strategies to use in these classes? We believe the question should be “what is the commonality in the student diversity”? Our answer is that we have had the students work together in groups of one-to-three and instruct each other. At first, we let the students self-select their groups. This gives each student a friend. If their grades were not acceptable, we would reassign them to another group. Research supports this strategy. Surveys using the Force Concept Inventory for the first semester of physics found that students typically doubled or tripled their learning gains in peer instruction courses (Hake, 1998; Crouch & Mazur, 2001; Lasry, Mazur & Watkins, 2008).

In Hattie’s 150 influences on achievement (2012, Appendix C), there are 69 influences (46 percent) with effect sizes above +0.40 (the expected 15 percent academic gain for one year of schooling). For the underachieving students, it is critical that we choose those strategies with high effect sizes (greater than +0.40). Otherwise, many of these students may never catch up academically, never pass the Texas STAAR tests and never graduate. However, we should not simplistically relate adjectives to the size of all effects. For example, Hattie (2012, pp. 13-14) noted a small effect requiring few resources might be more important than a larger effect size requiring high levels of resources. The effect size of reducing class size from 25-30 students to 15-20 students was (ES +0.22, +9 percent). But the effect size of teaching test-taking strategies was about (ES +0.27, +11 percent) in achievement and a +0.7 year gain in achievement.

Expanding Hattie's Research

To expand and broaden Hattie's 2012 research, we used Hattie's six major contributions to learning and conducted first a principal axes factor (PAF) analysis then a principal components analysis (PCA). These are both statistical procedures used to examine the correlations between variables and generate a factor structure based on those relationships. Principal components analysis is the default method commonly used in statistical packages and is used with considerable frequency when exploratory factor analysis (EFA) is performed (Russell, 2002; Thompson, 2004). The only difference between the two methods is that PCA uses one's (1.0's) in the diagonal of the correlation matrix. But in PAF factoring, the communality estimates on the diagonal of the correlation matrix are iteratively estimated until the iterations converge (Thompson, 2004, p. 53). We then conducted an initial PAF factor extraction (Kline, 1994, p. 75) with direct oblimin rotation and examined the factor correlations provided as part of the oblique rotation output. Since the factor correlations were less than 0.1, we used principal component analysis (PCA) with the scree test rotation. See Meyers, Gamst, and Guarino (2006, p. 502) for a discussion of these decision choices. Thus, we conducted a PCA analysis (Cattell, 1966) and parallel analysis (Horn, 1965). The rotated factor component matrix values follow: student (.965); home (.934); school (.911); teacher (-.837); curriculum (.988); and teaching (.540). Our analysis found three macro (composite) components. The number of components was confirmed by both a scree test (Cattell, 1966) and parallel analysis (Horn, 1965). The student (15%) and home (12%) combined to account for 32% of the total student achievement variance. The teacher (18%), curricula (17%), and teaching (17%) combined to account for 33% of the total variance, and the school accounted for 19%. The total variance accounted for was over 80%. This analysis showed that teachers (teacher, curricula,

and teaching) were the major contributor to student learning. These composite results have provided a much broader context for understanding and applying Hattie's 2012 research findings.

High-Stakes Testing

The last section of the presentation dealt with student high-stakes testing. For over a decade, the presenters have studied and applied the research literature pertaining to student achievement. This has borne very good results. For the years of TAKS testing, we carefully studied how the TAKS test was designed and scored. We talked with our students about the five TAKS science objectives and number of questions on each objective. We told our students the number of questions needed to pass the yearly science TAKS and gave them strategies for answering questions. We also gave all the students a summary sheet of their past scores, noting areas of strength or weakness. We also gave each student remedial work based on his/her previous test scores. We then shared our summary of all the previous year's test scores. We found, for example, if students answered 10 or more questions correctly on objective one (out of the 17 questions on the first objective), then 95 percent of our students passed science TAKS. If they answered less than 10 questions correctly on the first objective, they would have had to answer seven or more questions correctly on objective five to pass the TAKS test. This process provided information regarding the value of correct responses on the test.

The authors also had questions about the factor structure of the science TAKS test. Therefore, we conducted a SPSS factor analysis of our pool of TAKS scores. We shared our findings with the students and showed them the significance of the further clarification of the test structure. We also explained why students often failed the science TAKS test. Reasons included

class attendance, student's background, student's analytical skills, science vocabulary, language problems, and an inability to plan, organize and project one's life into the future (versus living only in the present). Some of the students in the classes had failed every science TAKS test they had ever taken. Other students had failed their last science TAKS test. We also knew the indicators of student dropout were poor attendance the freshman school year, failing grades the sophomore school year, and a reading level below the 20th percentile. Also see the Texas Education Agency's (2010) E0910 AT-RISK-INDICATOR-CODE document.

For 12th grade students, not passing TAKS meant they would not graduate with their senior class. We believe this increased the pressure on the students to pass the test and influenced their effort in class. Furthermore, this pressure was the reason some withdrew from school – many bored, angry and deliberately not learning. As we discussed interpersonal skills, we observed this had a positive effect on several students. This was related to the social-emotional aspect of achievement. Overall, many were the Myers-Briggs INFP profile students. Less than 15% were the ESTJ-type student.

The last year of TAKS testing, we were given our schools science department's TAKS test scores. My students outscored the science department at the 0.05 level of significance. My chemistry classes, scores reported by five groups (African American, Hispanic, Caucasian, economically disadvantaged, and special education), had higher scores ($M=80.40$, $SD=9.18$) than the overall science department ($M=64.0$, $SD=13.0$). The results of the independent samples t-test by group follow: $(t(8)=2.29, p=.05)$.

Following the end of the TAKS high-stakes testing, the State of Texas piloted a STAAR chemistry test. The following academic year, students took the one and only Texas' STAAR

chemistry test. Using strategies discussed in this document, for my on-level students there were more commended students on the STAAR test than any teacher teaching pre-AP, AP, or IB students at my high school. Using the students' State of Texas pilot STAAR test scores and their previous year's TAKS scores, the presenters published a logistic regression document:

“Predicting Student Success on the Texas Chemistry STAAR Test: A Logistic Regression Analysis” (Johnson, Johnson & Johnson, 2012). The ERIC document, with the SPSS source code, was also published in the ERIC document.

Item Response Theory

Item response theory (IRT) has become a favored methodological framework for developing assessments in education. The methodology has advantages over classical test theory (CTT): it focuses more on individual item analysis than CTT. Specifically IRT test development depends both on independently estimated item and ability parameter invariance. In IRT, a latent variable is fit to responses in a series of test items. IRT is not classical test theory based on the general linear model (GLM). See the ERIC Document Service No. ED 335 369 (Johnson, Ryan & Dixon, 1991).

Summary and Conclusion

In the presentation at the 2019 STAT Science Conference in Dallas, Texas, we looked inside the education black box and examined several important components of student academic success. We looked at why students failed and indicators of student dropouts. We also looked at our students' achievement and focused on teaching strategies to increase student achievement. We examined Robert Marzano's research and then discussed the research of John Hattie. Promoting the use of effect sizes in research has represented a major advancement in

understanding the components of high student achievement. We last discussed state testing and gave examples of how both the TAKS and STAAR tests were constructed, how our students scored on the Texas' state tests, and strategies we used to increase student achievement.

References

Ball, D.L. & Cohen, J.K. (1995, April). What does the educational system bring to learning a new pedagogy of reading or mathematics. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Cattell, R.A. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 30, 245-276.

Coleman, J. (1988). Social capital in the creation of human capital. *American Journal of Sociology*, 94, S95-S120.

Crouch, C.H. & Mazur, E. (2001). Peer Instruction: Ten years of experience and results. *American Journal of Physics* 69 (9): 970-977.

Ellis, P.D. (2016). *The Essential Guide to Effect Sizes*. Cambridge, United Kingdom: Cambridge University Press.

Flanagan, D. & Ortiz, S. (2001). *Essentials of Cross-Battery Assessment*. New York: John Wiley & Sons, Inc.

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics* 66, (1): 64-74.

Hattie, J.A.C. (1992). Measuring the effects of schooling. *Australian Journal of Education*, 36(1), 5-13.

Hattie, J.A.C. (October 2003). Teachers make a difference: What is the research evidence? Paper presented at the Australian Council for Educational Research Annual Conference on Building Teacher Quality, Melbourne.

Hattie, J.A.C. (2009). *Visible learning: A Synthesis of over 800-Meta Analyses Relating to Achievement*. London: Routledge.

Hattie, J.A.C. (2012). *Visible Learning for Teachers: Maximizing impact on learning*. London: Routledge.

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30, 179-185.

Johnson, W.L. & Johnson, A.B. (1995). Using SAS/PC for higher order factoring. *Educational and Psychological Measurement*, 55(3), 429-434.

Johnson, W.L. & Johnson, A.B. (2000). Using SPSS/PC for higher order factoring. *Educational and Psychological Measurement*, 60(4), 648-649.

Johnson, W.L. & Johnson, A.B. (2016). Student-centered learning: The new Texas teacher evaluation system. *Texas Study of Secondary Education*, 26 (1): 17-19.

Johnson, W.L. & Johnson, A.B. (2018). In search of the great schools: Closing the achievement divide. *Texas Study of Secondary Education*, 27 (2): 4-7.

Johnson, W.L., Johnson, A.B., & Johnson, J.W. (2012). Predicting Success on the Texas Chemistry STAAR Test: A Logistic Regression Analysis. (ERIC Document Reproduction Service No. ED 534 647).

Johnson, W.L., Ryan, J.M., & Dixon, P.N. (1991, April). A factorial and Rasch analysis of the Charles F. Kettering Ltd. School climate profile. Paper presented at the annual meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 335 369).

Kaiser, H.F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187-200.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. New York: Routledge.

Lasry, N., Mazur, E., & Watkins, J. (2008). Peer instruction: From Harvard to the two-year college. *American Journal of Physics*, 76 (11): 1066-1069.

Marzano, R.J., Marzano, J.S. & Pickering, D.J. (2003). *Classroom Management That Works: Research-based Strategies for Every Teacher*. Alexandria, VA: ASCD.

Marzano, R.J., Warrick, P. & Simms, J.A. (2014). *A handbook for high reliability schools: The next step in reform*. Bloomington, IN: Marzano Research Laboratory.

Matos, A. (2018, September 28). Texas top-ranked high schools don't prepare most kids for college, data shows. Retrieved July 26, 2020 from <https://Houstonchronicle.com/news/politics/texas/article/Texas>.

Myers, L.S., Gamst, G. & Guarino, A.J. (2006). *Applied Multivariate Research*. Thousand Oaks, CA: Sage Publications, Inc.

Peters, J. (2004). School environment: Putting staff No 1 connects students. Retrieved July 30, 2009, from www.aasa.org.

Russell, D.W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis. *Personality and Social Psychology Bulletin*, 28 (12): 1629-1646.

Spillane, J.P. & Thompson, C.L. (1997). Reconstructing conceptions of local capacity: The local education agency's capacity for ambitious instructional reform. *Educational Evaluation and Policy Analysis*, 19, 185-203.

Texas Education Agency (2010, July). E019 AT-RISK-INDICATOR-CODE. (<http://ritter.tea.state.tx.us?standards/1314/e0919.html>). Austin, TX.

Thompson, B. (2002). What future quantitative social science could look like: confidence intervals for effect sizes. *Educational Researcher*, 31(3): 25-32.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis*. Washington DC: American Psychological Association.

Thut, I.N. (1957). *The Story of Education*. New York: McGraw-Hill