

# TKES Research in Georgia Public Schools: Measuring Student Growth and Achievement

Carrie O'Bryant, Ed.D., James L. Pate, Ph.D., Lantry L. Brockmeier, Ph.D.,  
Barbara J. Radcliffe, Ph.D., Kathy Nobles, Ed.D., and Kelly C. Dreger, Ed.D.

of Department of Leadership, Technology, and Workforce Development  
Valdosta State University  
Valdosta, GA, USA  
jlpate@valdosta.edu

**Abstract—** Effective school systems implement evaluation tools to help students and teachers improve over time. Unfortunately, there is a lack of clarity on the appropriateness of current evaluation tools. A quantitative research study was conducted to determine the extent to which Georgia Teacher Keys Effectiveness System (TKES), academic setting, and percentage levels for students with economic disadvantages predicted student growth and achievement. Regression analyses and ANOVAs were used to analyze data for 4,000 educators in English Language Arts and Math. Results showed that TKES standards were predictive of growth and achievement for fourth graders and fifth graders. Economic disadvantage was a significant predictor for the 4th and 5th grades. The findings are of interest to educators and other stakeholders in K-12.

**Keywords—** evaluation; growth; achievement; school setting; standards; economic status

## I. INTRODUCTION

There are two well-known strategies for gathering useful information about student learning: assessment and evaluation. Assessment is generally described as an objective, systematic way to gather and review information about student learning. Evaluation, in contrast, is a subjective, problem-solving avenue for collecting information about students and judging the value of the data produced [1, 2]. Educational leaders and researchers generally agree that effective evaluation tools are critical for improving the teaching and learning processes [3-6]. There is, however, a lack of information available on how they reasonably align with research-based practices, student growth, and student achievement [7, 8].

One current evaluation tool that needs more in-depth empirical exploration is the Georgia Teacher Keys Effectiveness System (TKES). Originally piloted in 2012 as part of the Race to the Top Initiative, the TKES has a variety of factors for evaluation [9, 10]. One of these factors involve standards for the Teacher Assessment on Performance Standards (TAPS) [11]. These TKES standards are heavily based on research studies and practices developed

by Hattie [12]. There are many proponents of the TKES system who encourage its use as a valid, reliable, and integral part of addressing the diverse needs of students. It was designed to improve the professional development of Georgia teachers, which in turn improves opportunities for learning. There are just as many critics who strongly suggest that more evidence must be present to establish a significant impact between the system and student outcomes. It is possible for standardized evaluation tools to put more pressure on an already stressful situation concerning testing. It is not enough to enact new legislation and tools for success. Judgments need to be supported by facts and evidence-based practices.

With this in mind, a quantitative study was created to determine how well the TKES evaluation tool predicts student growth and achievement. Academic setting and economic disadvantage were also analyzed to determine their effects on growth and achievement. The overarching research questions were the following: (1) To what extent are summative scores on TKES standards significant predictors of teachers' mean scale scores and percentile levels from the Georgia Milestones Assessment System (GMAS)? and (2) Do significant differences exist in teachers' mean scale scores from the Georgia Milestones according to academic setting and percentage levels of economically disadvantaged (ED) students? The findings of the study help (a) increase awareness of how the TKES can link teacher behavior with professional outcomes and (b) provide understanding for educators, administrators, and researchers about the role TKES plays in outcomes related to student growth and achievement.

## II. LITERATURE REVIEW

Evaluation is a problem-solving activity that likely started as a tool for evolution, helping early humans with decisions on how to survive [1]. More formalized approaches to evaluation occurred over time. This includes diet regimen evaluation mentioned in the Bible's Book of Daniel and personnel evaluations that happened in 2200 B. C. [1,13]. Teacher evaluations were not deemed as a noteworthy pursuit until the 1700s, even though educational leaders' evaluations of teacher effectiveness started around the fifteenth century [3, 8]. Critical grade-based teacher

evaluations became essential parts of philosophies espoused by John Dewey and Frederick Taylor, particularly on the topic of public education [7]. Several major shifts in evaluation occurred between the 1930s to 1980s [7, 14, 15]. For instance, Goldhammer [14] developed a five-phase process of supervision that helped teachers with the reflection process. Hunter [15] created guidance for lessons, professional development, and supervisory conferences where evaluations for teachers and administrators played a key role. According to Marzano, Frontier, and Livingston [7], the key pioneers in modern evaluation methods were Glatthorn, McGreal, Glickman, and Danielson. Modern teacher evaluation models are often based on a combination of their recommendations. Their practices emphasized the incorporation of career-based goals, needs-based professional development, instructional improvement, and performance-level teacher evaluations [16-19].

#### A. Modern Teacher Evaluation Models

Although there are many avenues for evaluation available for school systems today, there is still debate as to which ones are necessary for academic growth. A variety of studies pertaining to educational evaluations have discovered positive associations concerning school improvement, evaluation, instruction, and achievement [20-25]. The value added model (VAM), observation-based model, Focused Teacher Evaluation Model [26], and Danielson's Model [27] are commonly used evaluation strategies, but more is desired from them in terms of practical use and significance. For instance, Baker et al. [20] found that VAMs lacked proper feedback mechanisms for evaluation, and academic ranking was not determined by how well teachers performed during the year. While they are based on formal assessments, more was needed to actually determine teacher effectiveness. Milanowski [28] adds to this by saying a combination of methods that may or may not include VAM was most appropriate for teacher evaluation, stressing more capabilities for teacher effectiveness and day-to-day improvements. Observations, walkthroughs, and artifacts would be components of any evaluation system created. Raudys [29] determined that observations allowed for reliable evaluation results via rubrics, but the quality and timing of the feedback may not accurately portray the day-to-day events that can happen during instructional periods. Marzano found through the use of the Focused Teacher Evaluation model with state testing that scores from systematic observations significantly predicted student growth [26]. Danielson's Model [27] is a modification on the rubric-based model created in 1996, and it addresses the Common Core Standards. It also includes a combination of collaboration and evidence-based communication not previously discussed, which would support the need for a combination of strategies.

#### B. The TKES Model

This study emphasizes the use of the TKES Model, which combines effective strategies used from the models already mentioned. The TKES generates a Teacher Effectiveness Measure (TEM) rating for teachers and administrators. This calculated measure is completed yearly. The TEM is based on three components within the system: the Teacher Assessment on Performance Standards (TAPS), Student Growth, and Professional Growth [5]. TAPS uses rubrics, observations, and artifacts to rate teachers according to 10 performance standards and 5 domains. The five domains are (1) Planning, (2) Instructional Delivery, (3) Assessment of and for Learning, (4) Learning Environment, and (5) Professionalism and Communication. The ten standards are Professional Knowledge (Standard 1), Instructional Planning (Standard 2), Instructional Strategies (Standard 3), Differentiated Instruction (Standard 4), Assessment Strategies (Standard 5), Assessment Uses (Standard 6), Positive Learning Environment (Standard 7), Academically Challenging Environment (Standard 8), Professionalism (Standard 9), and Communication (Standard 10). Figure 1 explains how the domains and standards are connected to one another. The standards are similar to the interactive ideas found within John Hattie's theory of high-impact strategies as well as his list of strategies that affect student achievement [12, 30, 31]. The strategies that he labeled as ones with high effect sizes (.60 or larger) are included in the TKES domains and standards [12, 32, 33].

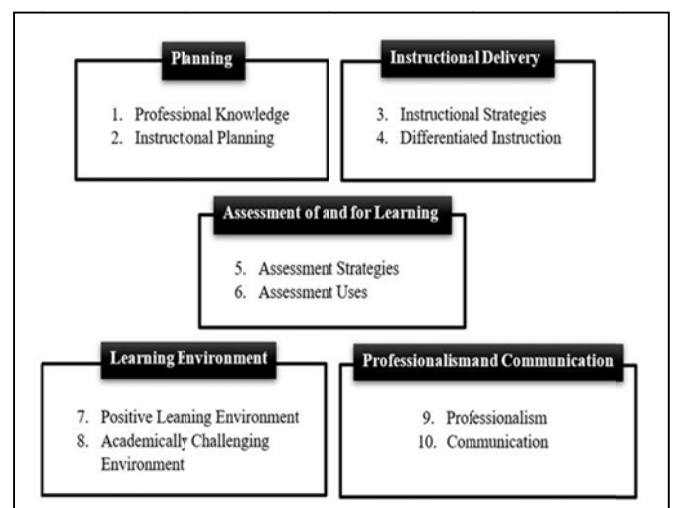
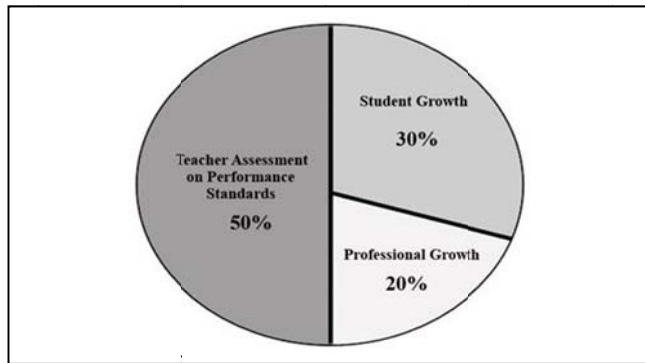


Fig. 1. Domains and Standards for TKES.



Components

Figure 2 illustrates the overall breakdown of the TEM components. TAPS accounts for 50% of the TEM. An additional 30% of the TEM is based on Student Growth, which is based on student percentiles that measure the progress of students throughout the year. The final 20% of the TEM is Professional Growth, which is a measure of how well a teacher met a plan or goal assigned to them based on their effectiveness and experience. The overall TEM score has four possible ratings: exemplary, proficient, needs development, and ineffective. This comprehensive rating is ultimately designed to help teachers and administrators reflect on their practices and make significant adjustments in order to promote growth at all levels of teaching and learning.

### III. METHODS

Permission was obtained from the university's institutional review board and the Georgia Department of Education (GaDOE) to conduct the study. GaDOE sent information about teachers and students from their database. The datasets received were for the 2021 and 2022 school years. The target population involved fourth and fifth grade teachers because all relevant score and percentile data were available for students who were at least in the fourth grade. GaDOE delivered datasets for a random sample of 4,000 teachers in fourth and fifth grade, specifically for the subjects of English Language Arts (ELA) and Math. All of the GaDOE information was placed into R statistical software for analysis. Regression modeling and ANOVA testing were used to answer the research questions. The independent variables within the datasets were TKES summative scores, levels of the percentages of students with economic disadvantage (ED), and academic setting. The summative scores were based on the four ratings for the TEM. There were two levels of ED students for the purposes of this study: (1) Schools with 0 to 75.56 percent of students classified as ED and (2) Schools with 75.57 to 100% of students classified as ED. Academic setting was either departmentalized or self-contained. A departmentalized setting required educators to teach ELA or Math with different groups of students each school day. The self-contained setting required educators to teach all subjects to a group of students each school day. The dependent variables were the student growth percentile (SGP) and mean scale

score (MSS) from the GMAS. The SGP is representative of student growth, and the MSS is representative of student achievement.

#### A. Validity and Reliability

The TKES evaluation process was validated using Kane's approach to validity, which is known as the Arguments-based approach [34, 35]. This approach requires an interpretive argument that contains four essential inferences: scoring, extrapolation, generalization, and implications. In terms of scoring, the TAPS process uses a rating system that is explicitly tied to TKES standards and domains. The TKES standards align with InTASC Model Core Teaching Standards and Learning Progressions for Teachers [36]. The GMAS was developed by the Georgia Legislature with the help of professional educators. Students in grades three to eight are tested on subject areas taught in schools today. Mean scale scores are generated that are then used to categorize students according to four different achievement levels: beginning learners, developing learners, proficient learners, and advanced proficiency [37]. For extrapolation, all results gathered from the TKES and GMAS are tied to performance-based standards and practical evidence of those standards. Teachers and students cannot be effective unless they demonstrate that effectiveness with evidence. Results from samples of teaching and learning are included as part of calculations for student growth and the TEM. The general, theoretical premise of TKES is that there are constructs used and defined for the following activities: planning, professional development, teacher evaluation, administrative evaluation, career-based decision-making, and intervention development [5, 9, 10]. The GMAS is designed to measure student progress and potential areas of improvement. The implications for both the TKES and GMAS are that they are informative measures for local, state, and federal policies concerning education. They provide educational stakeholders with critical information about teacher development and student progress.

Reliability for TKES is established with TAPS. A high ordinal alpha of .95 was calculated for TAPS. This calculation was created by the Georgia Center of Assessment [11], and it shows there is high simulated consistency between and among items. Internal consistency reliability for the GMAS was established with Cronbach's alpha, with overall reliabilities ranging from .88 to .94 [38]. Average reliabilities for the ELA area of the test range from .90 to .91 for students in grades 3 through 5. For the same grade levels, average reliability in the Math section is .93.

### IV. RESULTS

Ordinal logistic regression was used to determine the extent to which TKES standards predict SGP. Four assumptions about the data were met: (a) The dependent variable was binary during analysis, (b) independence is established between observations, (c) any existing correlations were less than .90, and (d) no independent variables were correlated during modeling procedures. There were two indicator variables constructed for SGP:  $SGP \geq 2$  and  $SGP \geq 4$ .

These levels were based on coefficients generated for SGP during regression procedures. General SGP ratings for teachers are provided in Table 1.

TABLE I. SGP RATINGS FOR TEACHERS BY GRADE LEVEL AND SUBJECT

Grade	Ineffective	Needs Development	Proficient	Exemplary	Mdn	M	SD
4 ELA	17 (1.21%)	160 (11.43%)	1135 (81.07%)	88 (6.29%)	50.38	50.46	46.91
4 Math	52 (3.71%)	211 (15.07%)	992 (70.86%)	145 (10.36%)	51.05	50.84	62.32
5 ELA	10 (0.71%)	148 (10.57%)	1165 (83.21%)	77 (5.50%)	50.24	50.63	43.03
5 Math	74 (5.29)	216 (15.43)	947 (64.64)	163 (11.64%)	49.78	50.16	67.96

Adapted from O'Bryant [8].

In fourth grade, Standard 4 and Standard 8 predicted SGP for ELA in a positive direction. A negative predictor/response association was found with Standards 9 and 3. Standards 1 and 3 positively predicted student growth in math. No predictors existed for fifth grade student growth.

Multiple linear regression models were used to determine the extent to which TKES standards predicted the MSS of students. The standards were coded into three levels: Level 2 (Needs Development), Level 3 (Proficient), and Level 4 (Exemplary). MSS was calculated as the average of all MSS generated for each subject-based class. MSS descriptives sorted by grade level and subject are provided in Table 2.

TABLE II. GMAS TEACHER'S MEAN SCALE SCORES (MSS) BY GRADE LEVEL AND SUBJECT

Grade	Mdn	M	SD	Skew	Kurtosis
4 ELA	507.78	508.56	29.39	0.79	0.00
4 Math	522.14	524.62	30.02	0.80	0.51
5 ELA	511.50	514.42	32.04	0.86	0.07
5 Math	509.80	514.85	32.80	0.81	1.06

Adapted from O'Bryant [8].

Assumptions about homoscedasticity (i.e., equal variation) of residuals and multicollinearity were met without transformations. As in ordinal regression, any correlations that existed before multiple regression analysis were less than .90, which ruled out a direct relationship between variables. Box-cox transformations were used to meet the normality assumption. Results showed that there were predictors that existed for student achievement in fourth-grade ELA ( $p < .05$ ): Standard 3 (exemplary), Standard 7 (exemplary), Standard 8 (exemplary), and Standard 10 (exemplary). The association was

positive for all standards except Standard 8, which had a negative impact. Fourth-grade math MSS had the following exemplary-level predictors: Standard 1, Standard 2, Standard 7, and Standard 8 ( $p < .05$ ). There were four predictors of student achievement for fifth-grade ELA and Math: Standard 1, Standard 4, Standard 7, Standard 8, and Standard 9. All were positive predictors at the exemplary level ( $p < .05$ ).

Factorial ANOVAs were used to determine significant differences in MSS according to academic setting and ED percentage levels. Data met the assumptions of observation independence, normality, and equal variances as required for parametric testing. Yeo-Johnson transformations were used to establish normality of data and homogeneity of variances. Academic setting and ED classification had two levels each as previously described, and MSS was a continuous interval variable for the analyses. ED levels were statistically significant for fourth and fifth grade ( $p < .05$ ), where schools with lower ED levels had higher student achievement. This was found for both ELA and Math scores. There were no consistent differences found for scores categorized by academic setting, but there was a significant difference found according to grade level and subject. In fourth grade math, departmentalized students significantly scored higher than self-contained students ( $p = .02$ ).

## V. DISCUSSION

Overall, the findings support the use of the TKES in the school system, but improvements need to be made so that achievement and growth can be consistently tied to the standards. Standards 1, 2, 4, 7, and 10 had positive associations with student progress. As teachers increase their proficiency in terms of professional knowledge, instructional planning, differentiated instruction, creating a positive learning environment, and communication, the chances for student progress (i.e., growth or achievement) increase. Unfortunately, the amount of impact varied according to grade and subject level. Only Standard 7 was a positive predictor for achievement in fourth-grade ELA, fourth-grade Math, fifth-grade ELA, and fifth grade Math. Standards 9, 3, and 8 had mixed results. Standard 8, which pertains to academically challenging environment, was a significant predictor for the different grade-subject combinations as well; however, it was not a positive predictor for all of them. This means that there are situations where a challenging environment is not appropriate. Teachers proficient in that area would need other skills and information to be able to meet the needs of students. Having the skills to create challenge, implement instructional strategies, and have professionalism in the classroom may not always match the practical needs of the job. The events that happen in the school system can change from day to day. Unexpected events could happen where teachers would have to know how to adapt and adjust their methods while keeping in mind the required skill set for teaching.

Surprisingly, Standards 5 and 6 were not significant predictors for the grades and subject areas studied. This points to the fact that assessment does not tell teachers and administrators everything that needs to be known about students. A combination of different strategies are needed in order to determine growth and achievement for a combination of different students. This idea is supported by Baker et al. [20], Milanowski [28], the improved version of Danielson's Model [27], and Raudys [29]. The strategies needed are not limited to academic assessments and evaluations. There are tools available, such as interest inventories, creative prompts, and personality quizzes that do not have to have an academic purpose. They could be social, emotional, recreational, or psychological in nature, but they may be able to provide information about students and teachers that academic-based tools currently lack.

Finally, academic setting did not show consistency in its results, but levels of economic disadvantage did. Economic disadvantage needs to be taken into account when looking at achievement scores. It is unclear what is required to further alleviate or mitigate circumstances concerning income, resource access, and affordable meals. Students who are at-risk in terms of behavior, health, or income need additional support, not just academically, but in all major areas of their lives. Their academic performance is just one piece of their complex situation.

#### A. *Limitations and Recommendations*

There are a few limitations within the study. One limitation is that the results only apply to Georgia teachers and students in fourth and fifth grade, particularly within the areas of ELA and Math. It is recommended that more research needs to be completed that expands the information to more locations, grade levels, and subject areas. The second limitation pertains to the fact that evaluation tools are subjective in nature. Having different, reliable raters for TKES evaluations or including other rating tools within a study can better address any inconsistencies or biases that may arise as a result of having evaluations rather than assessments. A survey about evaluation tools would provide additional feedback about how administrators and teachers utilize evaluation strategies as a whole. Thirdly, the SGPs used for the study were based on information from 2018 and 2019 GMAS testing. Due to the COVID-19 pandemic, there were no GMAS scores available for 2020, which was the most recent year for the datasets received. Some parents also decide to not give consent for their children to take the test. A future study can be completed that replicates the procedures in this one while also using updated data for GMAS. Finally, there are factors outside of teachers' control that may influence student achievement and growth. More studies need to be created that explore additional factors, such as time constraints, resource access, incentive use, student behavior, parental involvement, socioeconomic status, health, attendance, and unique life events. Teachers need to address these factors

during instruction, connecting their lessons to practical learning experiences and events that students can benefit from within and outside of the classroom setting. For example, a math teacher could have students look at nutrition labels in order to learn how to calculate the number of calories in servings. This connects a health component to their academic learning that they could use throughout their lives.

#### VI. CONCLUSION

A quantitative study was completed about the Georgia TKES, academic setting, and economic disadvantage within public schools. Specifically, it addressed to what extent TKES standards predicted student achievement and growth as defined by the GMAS. It also determined if significant differences in GMAS scores existed according to academic setting and economic disadvantage. It is clear that some TKES standards are able to predict growth and achievement in fourth and fifth grade subjects, which makes those areas of the evaluation system relevant for achievement and growth outcomes today. There is a lack of consistency among the predictions, but the standards that showed up most frequently were Academically Challenging Environment (Standard 8) and Positive Learning Environment (Standard 7). The most positive and frequent impact was found with Positive Learning Environment (Standard 7). While academic setting only affected fourth grade math, the economic disadvantage designation consistently predicted outcomes for students in fourth and fifth grades within both ELA and Math. More research and practice needs to address how TKES can be better aligned to students' needs, including but not limited to their academic needs. Evaluating teachers is important for the development of effective educators, but educational policies and teaching practices will decide the extent to which TKES Standards can fully align with student growth and achievement.

#### ACKNOWLEDGMENT

All contributors work in the state of Georgia to educate students in need. They have a diverse range of backgrounds and experiences. All are interested in research within the areas of curriculum, instruction, leadership, teaching, learning, and technology use.

#### REFERENCES

- [1] M.C. Alkin and J.A. King. (2016). "The historical development of evaluation use." *American Journal of Evaluation*, 37(4), 568-579. <https://escholarship.org/uc/item/2qn5x84t>
- [2] L. Anderson, A. Butler, A. Palmiter, and E. Arcaira. (2016). Study of emerging teacher evaluation systems. <https://www2.ed.gov/rschstat/eval/teaching/emerging-teacher-evaluation/report.pdf>
- [3] C. Barrette, E. Morton, and A. Tozcu. (1995). "An overview of teacher evaluation." *SLAT Student Association*, 3(1), 36-48.

- [4] C. Carbough, R. Marzano, and M. Toth. (2017). The Marzano focused teacher evaluation model. <https://www.learningsciences.com/wp/wp-content/uploads/2017/06/Focus-Eval-Model-Overview-2017.pdf>
- [5] Georgia Department of Education (2020). Georgia's Teacher Keys Effectiveness System: Implementation handbook. <https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TKES%20LKES%20Documents/2020-2021/TKES%20Handbook%20Final%202020.2021.pdf>
- [6] T. Toch and R. Rothman. (2008). *Rush to judgment: Teacher evaluation in public education*. Washington, DC: Education Sector.
- [7] R.J. Marzano, T. Frontier, and D. Livingston. (2011). *Effective supervision: Supporting the art and science of teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- [8] C.E. O'Bryant. (2022). *Teacher evaluation and other implications of student growth and achievement in public schools in Georgia* (Doctoral dissertation). <https://vtext.valdosta.edu/xmlui/handle/10428/5434>
- [9] J.D. Barge. (2012). Overview to the 2012 TKES/LKES pilot evaluation report. [https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/Pilot%20Report\\_Overview%20and%20Report%20Combined%201-10-13.pdf](https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/Pilot%20Report_Overview%20and%20Report%20Combined%201-10-13.pdf)
- [10] Georgia Department of Education. (2012). *Teacher Keys Effectiveness System: Fact sheets*. <https://www.gadoe.org/School-Improvement/Teacher-and-Leader-Effectiveness/Documents/TKES%20Fact%20%20Sheets%207-11-2012.pdf>
- [11] Georgia Center for Assessment. (2014). *Assessing the validity and reliability of the Teacher Keys Effectiveness System (TKES) and the Leader Keys Effectiveness System (LKES) of the Georgia Department of Education*. [https://gaedassessment.files.wordpress.com/2018/05/tkes-lkes\\_reliability-validity\\_report\\_v2-student-perception-surveys-included.pdf](https://gaedassessment.files.wordpress.com/2018/05/tkes-lkes_reliability-validity_report_v2-student-perception-surveys-included.pdf)
- [12] J. Hattie. (2008). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*; Abingdon, Oxon: Routledge.
- [13] M.Q. Patton. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- [14] R. Goldhammer. (1969). *Clinical supervision: Special methods for the supervision of teachers*. New York, NY: Holt, Rinehart & Winston.
- [15] M. Hunter. (1980). "Six types of supervisory conferences." *Educational Leadership*, 37(5), 408-412.
- [16] C. Danielson. (1996). *Enhancing professional practice: A framework for teaching*. Alexandria, VA: Association for Supervision and Curriculum Development.
- [17] A. Glatthorn. (1984). *Differentiated supervision*. Alexandria, VA: Association for Supervision and Curriculum Development.
- [18] C. Glickman. (1985). *Development as the aim of instructional supervision*. <https://files.eric.ed.gov/fulltext/ED263655.pdf>
- [19] T. McGreal. (1983). *Successful teacher evaluation*. Alexandria, VA: Association for Supervision and Curriculum Development.
- [20] E. Baker, P. Barton, L. Darling-Hammond, E. Haertel, H. Ladd, R. Linn, D. Ravitch, R. Rothstein, R. Shavelson, and L. Shepard. (2010). *Problems with the use of student test scores to evaluate teachers*. EPI briefing paper #278. Washington, D.C.: Economic Policy Institute. <http://www.epi.org/>
- [21] M. Cogan. (1973). *Clinical supervision*. Boston, MA: Houghton Mifflin.
- [22] S. Cosner, S. M. Kimball, E. Barkowski, B. Carl, and C. Jones. (2015). "Principal roles, work demands, and supports needed to implement new teacher evaluation." *Mid-Western Educational Researcher*, 27(1), 76-95.
- [23] M.L. Derrington and J.W. Campbell. (2015). "Implementing new teacher evaluation systems: Principals' concerns and supervisor support." *Journal of Educational Change*, 16(3), 305-326. <https://doi.org/10.1007/s10833-015-9244-6>
- [24] H. Putman, E. Ross, and K. Walsh. (2018). *Making a difference: Six places where teacher evaluation systems are getting results*. <https://www.nctq.org/publications/Making-a-Difference>
- [25] D. Weisberg, S. Sexton, J. Mulhern, and D. Keeling. (2009). *The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness*. Brooklyn, NY: Grantee.
- [26] Marzano Center. (2018). *The Marzano focused teacher evaluation model: A focused, scientific-behavioral evaluation model for standards-based classrooms*. <https://www.marzano-center.com/wp-content/uploads/sites/4/2018/10/MC06-14-FTEM-White-Paper-1-16-18-Digital-4.pdf>
- [27] C. Danielson. (2014). *The framework for teaching evaluation instrument 2013 edition* [Ebook]. <http://www.loccsd.ca/~div15/wp-content/uploads/2015/09/2013-framework-for-teaching-evaluation-instrument.pdf>
- [28] A. Milanowski. (2011). "Strategic measures of teacher performance." *The Phi Delta Kappan*, 92(7), 19-25.
- [29] J. Raudys. (2018). *4 teacher evaluation models to use (with examples!)*. Retrieved from

<https://www.prodigygame.com/main-en/blog/teacher-evaluation>

[30] D. Fisher, N. Frey, and J. Hattie. (2016). Visible learning for literacy, grades K-12: Implementing the practices that work best to accelerate student learning. Thousand Oaks: CA: Corwin Press.

[31] E. Terhart. (2011). "Has John Hattie really found the holy grail of research on teaching? An extended review of Visible Learning." *Journal of Curriculum Studies*, 43(3), 425-438.

[32] J. Hattie, D. Fisher, N. Frey, L.M. Gojak, S.D. Moore, and W. Mellman. (2016). Visible learning for mathematics, grades K-12: What works best to optimize student learning. Thousand Oaks, CA: Corwin Press.

[33] Visible Learning. (2017). Hattie ranking: 250 + influences on student achievement. <https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/>

[34] M. Kane. (2009). Validating the interpretations and uses of test scores. The concept of validity: Revisions, new directions, and applications (1-73) Charlotte, NC: Information Age Publishing, Inc.

[35] M. Kane. (2013). "Validating the interpretations and uses of test scores." *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>

[36] T. Elder, J. Wang, and S. Cramer. (2015). An evaluation of the validity and reliability of the Intern Keys Assessment. <https://coehp.columbusstate.edu/docs/accreditation/an-evaluation-of-the-validity-and-reliability-of-the-intern-keys-assessment.pdf>

[37] Georgia Department of Education. (2023a). Understanding the Georgia Milestones Achievement Levels. [https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/achievement\\_levels.aspx](https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Pages/achievement_levels.aspx)

[38] Georgia Department of Education. (2023b). Validity and reliability for the 2021-2022 Georgia Milestones Assessment System. [https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Milestones/Technical\\_Documents/2021-22\\_Georgia\\_Milestones\\_Validity\\_and\\_Reliability\\_Brief.pdf](https://www.gadoe.org/Curriculum-Instruction-and-Assessment/Assessment/Documents/Milestones/Technical_Documents/2021-22_Georgia_Milestones_Validity_and_Reliability_Brief.pdf)