ISSN: 2475-5435

Review Article

International Journal of Psychiatry

Relationship between the Adoption of BARS Teacher Evaluation Model and its Impact on Student Performance

Richard Karnia*

Graduate Psychology Department, Purdue University Global

***Corresponding Author**
Richard Karnia, Graduate Psychology Department, Purdue University Global.

Submitted: 2024, Jul 18; **Accepted**: 2024, Aug 09; **Published**: 2024, Aug 20

Citation: Karnia, R. (2024). Relationship between the Adoption of BARS Teacher Evaluation Model and its Impact on Student Performance. *Int J Psychiatry, 9*(3), 01-18.

Abstract

There currently is a mismatch in the outcomes in education between teacher performance evaluation models and student outcomes. Data from Illinois State Board of Education shows teachers in one district have a 99.7% of teachers meeting excellent or proficient standards, while only 13% of 11th graders can read at grade level. The research will identify trends in performance evaluation and how the application of theories of industrial-organizational psychology can play in shaping and improving current teacher evaluation models. Professional development and evaluations should be linked to quantitative data which does not exist current evaluation models. The development of behavioral anchored rating scale would be a stronger evaluation model for teacher versus existing graphic rating scales. This should then lead to the development for a BARS model tied to each subject matter to increase the validity and reliability in teacher evaluations. Current teacher evaluation models do not properly assess teacher's knowledge, skills, abilities and other characteristics, and the implementation of a performance evaluation model for teachers based on a Behavior Anchored Rating Scale will lead to improvement is student outcomes.

I. Introduction

Performance assessment is a process of developing and organizing a formal evaluation process of individual employees, with a clear framework for evaluation that is designed to optimize the quality of work, identify poor performers, and develop individual worker's abilities (Kiplangat & Simon, 2021). Performance management plays a role in strategic human resources management and its goal is to developing a process of measuring, communicating, and managing employee performance to meet an organization's goal [1]. The effort of appraisals methods should be part of the realization of the members of an organization and the expectation is that performance assessment will have a positive effective on achievement (Kiplangat & Simon 2021).

When looking at current teacher evaluation methods, there is a mismatch between the performance evaluation measurement indicating high/successful evaluation scores, but low level of organizational success which would be student achievement. When looking at teacher performance levels and student achievement data, pre Covid-19 school closures, an example of the mismatch performance is in Decatur Illinois where the Illinois State Board of Education reports identifies in 2018 that 99.7% of teacher evaluations are excellent or proficient (highest rating), but student (organizational) achievement shows only 13% of students can read at grade level, 11% can do math at grade level, and only 74% of students graduate [2]. The purpose of any appraisal method is to provide employees with feedback about job performance and areas of deficient performance, and current teacher evaluation models appear to be subject to rating errors such as halo errors, central tendency errors, and leniency error which contaminate the ratings and performance scores and expected results of perceived scores [3]. A way to address the deficiencies is current teacher evaluation methods which use a graphic rating scale, is to develop an evaluation method based on a Behavior Anchored Rating Scale. A BARS provides concrete examples of different level which address the existing deficiencies and avoids being subject for a rater to decide what is above average [4].

2. Literature Review

The literature review will focus on building a foundation for understand what performance evaluations are and the role they play in professional development using principles of research in industrial-organizational psychology. The literature review is a review of 22 different areas of research to give the reader a broad understanding of the following topics: the purpose of performance evaluations and appraisals, role of professional development in

education and current teacher performance evaluation models, drawbacks and challenges of using graphic rating scales in performance evaluations, and the effectiveness and purpose of using a behavior anchored evaluation model. The studies are an evaluation of quantitative research and evidence of existing evaluation models, the strength and weakness of various types of performance evaluations, and rationale with evidence to support the transition towards the development of the BARS model as a new form of teacher evaluations to improve student performance.

3. Purpose of Performance Evaluations and Appraisals

Performance is the action and behaviors of individuals relevant to an organization's goals, measure an individual's proficiencies, effectiveness, productivity, declarative knowledge, procedural knowledge, and motivation [5]. John Campbell states that there are eight performance components which can be found in every job including job-specific task proficiency, non-job-specific task proficiency, written and oral communication proficiency, demonstrating effort, maintaining personal discipline, facilitating peer and team performance, supervision/leadership, and management/administration [5]. Performance appraisals are the process of systematically evaluating employees across various performance dimensions and are valuable to provide feedback to employees and management, but are meaningless if they are accompanied by ineffective feedback back system [6]. Performance appraisals and evaluations are used for a variety of reasons including feedback to employees, self-development, rewards system, personal decisions, and training and development [6]. Performance appraisal is an evaluation of employee performance and part of the process of performance management which measures, communicates, and manages employee performance aligned with organizational goals [1].

A performance evaluation should be an objective performance that uses a predetermined criteria to measure employee and share the information with them [1]. Performance appraisals should include meetings that give employees feedback in a one-on-one meeting in a regular basis, plays a role in employee development and problem solving, is used to support organizational decision making, gathers high quality data, and serve a legal purpose to protect both employee and organizational rights [1]. When creating a performance management system, the design of an assessment tool should include the purpose and desired outcomes of the performance appraisal, defining performance, determining the specific performance criteria, chose the rating method, chose the source of the performance information, and deciding how closely the rating should be tied to compensation [1]. The performance appraisal has two components, a development purpose and an administrative purpose. The developmental purpose the process of providing employees with feedback for the purpose of improving their performance [1]. The administrative purpose plays a role in organizational decisions and making decisions on employees pay, bonuses, training, and promotions [1].

4. Role of Professional Development in Education and Current Teacher Performance Evaluation Models

The focus of role professional development plays in education it to enhance teachers' use of evidence-based practices to improve instructional strategies [7]. One study conducted by Reddy was focused on 2195 students and 106 teachers and to see how the implementation of professional development plays in the improvement in the quality of instruction and student academic engagement [7]. Evidence has shown that professional development requires performance feedback in the authentic learning environment, and teachers need sustained skill specific support to keep from reverting to prior, less effective practices [7]. The research indicates that providing teachers with an effective and efficient coaching intervention will enhance teacher effectiveness and higher student achievement [7]. This study illustrates the need to have individualized professional development plans, focused on direct measurable goals that can foster skill development and mastery which is exactly the purpose of implementing the subject specific BARS model versus existing graphic rating scales that tend to vague and lack clear specific goal development.

Studies have been designed to show the effectiveness of teacher professional development and its relation to enhancing teacher knowledge which would then improve student achievement [8]. Professional development in education is often done by outsiders from the school over a single day, and these individuals lack an understanding of the school, curriculum, and community and is often ineffective [8]. Student achievement is directly linked to capacity of the individual teacher and the impact of a teacher knowledge base on student test performance has been shown to show statistically significant correlations in multiple states [8]. The methodology of the study by Gupta and Lee was to determine how well teachers mastered content skills during professional development, the extent of the implementation of the content and skills covered in the training, and the impact on student performance measures following professional development training [8]. Data was gathered by conducting surveys of the teachers, observations of classrooms of teachers in the study, and gathering of test results from students. The data on student performance on the STAR assessment found small gains, but it the group did not keep pace with the progress of the norm group [8]. There were found to be modest gains in writing scores, significant gains in reading scores, and largest gains were in SOL English scores [8]. The limitations is that the model lacked for an evaluation of the level mastery of the teachers before the professional development, and the in the future any professional development should take into account the background and experience difference in the teachers to determine the impact of the training. The goal of the study was to find a correlation between teacher professional development, and the association to student learning. While the evidence may not have supported the premise, this is an indication that the need exists for a better performance evaluation model to identify individual strength and weaknesses teachers tailored to identify a teacher's mastery of content knowledge and skills. With the development of

subject matter-based BARS each teacher would have individually tailored professional development plan to improve their mastery and in turn increase student performance.

When looking at teaching evaluation models, they are often reviewed by an individual with an emphasis on pedagogical and not subject matter expertise, and there are very few measurements created by SME leading to a gap for human resources to identify principles of performance appraisals and performance management [9]. The purpose of developing a teacher evaluation based on performance appraisal techniques will prove a continuous process of measuring, identifying, and developing an individual teacher with goals of producing feedback for developmental purposes and data driven decisions [9]. A performance evaluation tool should be able to have a clear criterion to measure effective and ineffective performance and identify all employee (teachers) major job responsibilities as determined by a SME [9]. Accurate data can be collected and matched to a job analysis to prioritize the KSAO's need to execute the job which has been challenging in education due to lack of objective clarity on defining and measuring elements of teaching performance [9]. One model that can be used is a BARS which when created effectively can define performance in clear behavior terms versus other rating scales with qualitative anchors that to support consistent reliability in ratings. The implementation of a BARS model can be effective as it involves SME and stakeholders to create critical incidents that a key to the dimensions of the job and levels of performance, focuses on job related functions, have a high level of reliability and validity, are legally defensible, and offer clear feedback to employees for growth and professional development [9].

In 2009 many schools across the United States as part of the American Recovery and Reinvestment Act decided to adopt performance-based evaluation models to receive Race to the Top Grants (Tanner & Miller, 2021). The policy required evaluation methods to be linked to measurable standards focused on multiple factors of teacher effectiveness, and many chose Charlotte Danielson's Framework for Teaching (Tanner & Miller, 2021). Research was conducted by the University of Chicago evaluate the Chicago Public Schools (CPS) adoption of the Excellence in Teacher model using the Danielson Model Framework for Teaching with the goal of improving teaching practices, and improve the correlation between teacher performance ratings and student outcomes [10]. Prior to the study Chicago had 93% of teachers were in the top two tiers of performance evaluations and only .3% were unsatisfactory [10]. In 2023, 86% of teachers in Chicago Public schools were rated excellent or proficient (top two categories of evaluations), while on the Illinois Assessment of Readiness only 26 percent of students met proficiency in ELA and 18% of students in math [2]. In 2009 a different assessment was used, the Illinois Standards Achievement test, 67.5% of students were meeting or exceeding standards on state testing [11]. Comparing data across states is difficult as students and states use different forms of assessment, have changed the type of assessment (ACT, SAT, or other achievement tests).

The Danielson Model and Framework for Teaching tried to demonstrate the strength of the model by demonstrating the validity and reliability of the assessment. The use of the data was based on the identify the validity between the evaluation model and the use of a value-added indicator, which was shown by Sean Corcoran of having a large margin of error in teacher ranking and student performance [12]. The value-added data can cause an in the ranking of teachers from being evaluations to fluctuate from the 43rd percentile to actually being either in the 15th percentile to the 71st percentile [12]. The Danielson Model also lacks reliability, as reliability between different raters were also very low 52% of principals gave distinguished ratings, while only 24% of trained observers gave a distinguished rating [10]. The leniency error in performance evaluations were a result of the name/labeling of measures on the rating scale and principles felt the need to give distinguished ratings to preserve the relationship with teachers who have previously received high ratings [10]..

5. Drawbacks and Challenges of Using Graphic Rating Scales in Performance Evaluations

The Danielson model used a graphic rating scale that only used four points of measurement: unsatisfactory, basic, proficient, and distinguished, but evidence in the development of rating scales state there should be no more than 10 points but you should also not go below 5 points [10,13]. If using only a 5-point scale, in the case of the Danielson model 4 points, you will often exclude the end points so you have only 2-3 usable points. A scale should consist of 7-10 points to choose from even if raters are avoiding the extreme values [13]. The nature of the category labels are also important and can affect results, the language needs to be consistent with the wording of the idea being evaluated, and a mismatch between the wording and labels will after the raters ability to understand what they think was meant and a lower reliability [13]. The ratings on the scale should be able to be measured quantitatively each alternative measured KSAO should be equally spaced along a scale of measurement [13].

A rating scale as a form completed by an evaluator (a rater, judge, or examiner) to make a judgment of relative standing with regard to a specified variable or list of variables [14]. A rating scale is often subject to rating errors which is a numerical or verbal judgment from the intentional or unintentional misuse of a rating scale [14]. Thus, for example rating scales are subject to a leniency error where the rater is lenient in scoring and tends to rate most employees highly [1]. The other extreme can also take place of a severity error in which the rater tends to rate all individuals with a low performance score [1]. As noted earlier, rating scales should include at least 5 points but often scaled better from 7-10 behaviors to avoid the central tendency error as raters tend to exhibit a reluctance to give ratings at high or low ends of the rating scale and tend to cluster in the middle of the rating scale [14]. Other factors that can influence and effect accurate quantitative

data gathered from a graphic rating scale is the halo error in which one or two positive dimensions observed by the rater influences a positive perception on all other ratings, and the horns error which giving continuous negative ratings on all dimensions of an evaluation if one dimension was viewed negatively [1]. The Halo effect occurs when managers of overly positive view of employees, and can lead to impact the objectivity of reviewers and mangers will give employee high ratings and fail to recognize areas needing improvement [6].

6. The Effectiveness and Purpose of Using a Behavior Anchored Evaluation Model

A more effective measure to evaluate teacher performance is a model based on a Behaviorally Anchored Rating Scales or BARS. A BARS was originally designed to assess skills that research has shown to be important int obtaining success while performing many types and levels of jobs [4]. The skills that are assessed include the behavioral expression of personality, motivational, attitudinal, and self-regulatory constructs that are deemed critical to job performance [4]. It can be stated that the skills BARS measures are psychosocial, socioemotional, noncognitive, or personality in nature [4].

BARS was created with the intention to address the deficiencies of existing Graphical rating scales [4]. GRS scales have been criticized for being vague or ambiguous and an invalid variance can occur as a consequence of a rater's idiosyncratic interpretations of the evaluative dimension that causes different supervisors to have different rating interpretations of GRS dimensions [4]. This will lead different supervisors to have different interpretations of what a measurement constitutes and lead them to rate the same employee differently even if they observed the identical sample of employees' behaviors [4]. BARS was created to address the deficiencies of the prior GRS and can provide concrete behavior examples of different levels of performance which will give a rater explicit standard when evaluating an employee's performance [4]. BARS provides a behavioral exemplar for the rater of a job performance/skill versus the rater having to interpret the said performance/skill [4]. BARS was originally developed by Smith and Kendall with a sequence of seven steps which are writing critical incidents, develop performance dimensions, recheck and retranslation, scale the critical incidents by computing the mean and standard deviation, and then develop a final instrument [15].

BARS was designed for evaluating the performance of a specific job or job within a job family and can contribute to goal accomplishment in virtually any organization [4]. One key component of a BARS is that is based on rational consideration that reduces construct-irrelevant variances in performance ratings with an emphasis on specific concrete observable behaviors connected to anchoring evaluative continuums [4]. These concrete measurements increase the validity of the rating scale. Since the BARS and its components can be traced to SMEs and it has been argued that the elements composing BARS is job experts' judgements, they will automatically possess job relevance [4].

When evaluating BARS, it is important to look at and understand the validity and reliability of the evaluation tool. There is a high level of validity to a BARS evaluation because the job holders and supervisors who know the job develop the behavior description and this expertise results in construct validity (McCoy, 2023). What this means is the evaluation matches operational definition and represents the actual behaviors of an employee's role within an organization [16]. The reliability of BARS is also very because any person or supervisor should score the same for employees as it is not subjective but is connected to a scale or behaviors/actions of the employee. These measured behaviors are consistent as an individual rater will focus on the employee's job performance, and not on other outside factors such as personality, because these behavioral statements are straightforward and there is little variance regardless of who is the assessed and assessor [16].

Many organizations should use BARS because behaviors are well defined and both managers and employees can understand them without extensive training or explanations [16]. When using BARS, it also creates a mutual understanding on what is being reviewed and opportunities for improvement [16]. BARS are also individualized as each position within an organization will have a unique set of role related behaviors [16]. There are also several disadvantages of using BARS that focus on the complexity and cost of creating.

Further evidence for the implementation of a BARS based model for teacher evaluations is the measure the effectiveness of BARS used outside of education. Research by Pazos found the use of a Behavior Anchored Rating scale known as the Comprehensive Assessment of Team Members effectiveness and how this model was used between individuals working on a project to evaluate knowledge, skills, abilities and found it to that the mode that was used to have a high level of validity in measuring contributions to the team's work, interaction with the team members, keeping the team on track, expecting quality, and having relevant KSA's [17]. The research shows that by using the CATME a type of BARS there was improvement throughout the project and an increase in their own and their peers' abilities to accomplish the task [17]. In Nursing, Reed found evidence that a BARS assessment tool that could lead to measure behaviors and that a learning process that can be quantitatively used in debriefing of nurses [18]. A BARS was used to conducted evaluations to address deficiency that were found in graphical based rating performance scales which are too vague or ambiguous [18]. A BARS assessment model provides concrete examples that raters can use while assessing/performing an evaluation and emphasize specific, concrete, observable behaviors to be judged [18]. As a result, it is noted that BARS evaluations exhibit less bias than other scales and can be used to measure behavior aspects of a learning process [18]. The evidence indicates a BARS assessment, and the data gathered is used as a debriefing tool, it has the potential to contribute to the participant

teaching [18].

When looking at workplace-based assessment or performance evaluation models, there is a lower level of inter and intra rater reliability often associated with lack of training or the effectiveness of the rater training [19]. A transition towards a behavioral anchored rating scale is being used to show a progression toward competence as they are more concrete, transparent, and cognitively aligned which provides specific actionable feedback over numerical ratings [19]. The methods of the study were to conduct a single blind controlled trial of longitudinal rater training interventions and determine the impact on reliability of rating scores [19]. The process involved 2 in person 3-hour workshops along with the gathering of baseline data before and after the training/intervention [19]. Data was analyzed using a paired T-test of the intervention/training group versus a control group and when comparing scores at baseline there was no significant difference between the groups, but after training the intervention group had higher mean of accurately identifying behaviors (3.27 vs 2/29, $P < .001$) [19]. This shows that the rater training did have a positive impact on the quality of the assessment by the rater and the accuracy of the observations. When looking at the process of implementing a BARS model for education, the literature shows more evidence to support the rater training associated with the observational model. The raters need to have an extensive training and with clear content and guidelines. The evidence presented also supports the adoption of a BARS versus an ordinal scale will associate specific workplace activities that will demonstrate work progression towards gaining competence [19].

The goal of developing a BARS teacher performance appraisal is to create an assessment of an employee's (teachers) abilities which will measure a parameter of employee attitudes and behaviors on a measurable scale with anchored values that will show performance [20]. Bars models are created to define and assess behavioral parameters that are in decline, need improvement, and reduce the subjectivity of the rater and produces objective decisions based on the assessment [20]. The use of a graphic rating scale is easy to design, but can be objective based on the rater [20]. The use of BARS appraisal combines the critical incident rating and clearly defines with evidence/examples good and bad performance with a behavior for each dimension on a scale [20]. The use of a BARS method will allow an employer to measure attitudes, behaviors, and assessment parameters according to the unique process of each company on a measurement scale that can be objectively carried out [20].

Other evidence of effective use of BARS assessments shows the use of a BARS assessment to analyze non-technical skills, but observable behaviors including decision making, planning, situational awareness, social, and interpersonal skills [21]. The study was conducted by video review of 30 clerkship year medical students in simulation scenarios that were developed that were selected from the pediatric clerkship curriculum [21]. The participants were observed using an ordinals scale form from 1-9 based on observed behaviors [21]. The raters making the observations were given a brief BARS training and a matrix with examples of behaviors to be observed with an associated score. The reliability of inter-rater reliability of the assessment was .667 which shows insufficient evidence to say the BARS assessment has a high level of inter-rater reliability [21]. It was also noted that the BARS was just used as an assessment tool, but not a model for growth/professional development for the students. The information gathered from each BARS assessment should be used as a guide for improvement and shared with the medical students so they can develop the NTS that were not observed. The low inter-rater reliability scores may also be due to lack of training, as the raters only received 2 hours of training and familiarity with the scale [21]. Most BARS assessment requires the raters to be part of the development of the critical incidents for the model to be effective which may in turn lead to a higher level of reliability. The key takeaway from this research is the how important rater training and the development of critical incidents is needed for the proper application and use of BARS as a performance evaluation model. It was noted in the research that one flaw was that there was a need for more extensive training on the BARS and evaluation of raters learning curve should have been conducted would benefit in future studies [21].

7. Conclusion

It takes time and money to develop, review, and calibrate every performance level for each behavior which is also expensive as will require experts to create such as an Industrial Organizational Psychologists or Consulting firm to create [16]. Assessing an employee also does take time as it may include 60 or more behaviors being measured and requires frequent updates as skills are constantly changing [16]. Many more companies will hopefully transition to BARS as it can improve performance, engagement, employee retention and a very effective tool to reduce bias in evaluation and disparate impact [16]. The hypothesis is to gather quantitative data that will show the development and implementation of Behavior Anchored Rating scale for a new teacher evaluation model, tailored to specific critical incidents associated with individual secondary school subject matter, will result in a greater relationship between teacher performance evaluations and student success outcomes.

8. Method

8.1 Research Design

The focus of my research is to determine the effects and impacts of a performance evaluation method for teachers, and if its implementation will have an impact on student performance/outcomes. The goal would be to find an schools in a district with similar student demographic and have a school implement the performance evaluation model and the second school use the existing model and see if there is a difference between the experimental (new model) and control groups (existing model) on student achievement and performance on standardized

assessments.

The research design method that I choose was a correlational study as the goal is to determine if a changes in one variable are associated with changes in another [22]. This is an effective research method as it is trying to demonstrate that the cause happens before the effect which will establish temporal precedence [22]. This goal of the experiment will be to demonstrate there is a relationship between the independent variable (performance evaluation model) and the dependent variable (student achievement) which is the covariation of cause and effect [22]. By comparing between two similar schools and establishing pre and post test scores for students in each school, the use of the control group to compare will demonstrate that there were no other variable responsible for the effect [22].

Another research design would be a longitudinal design over the same school district and teachers to see the long term effects the implementation of a new teacher evaluation model. A longitudinal study follows the same group of participants over an extended period [13]. The problem of using this method is that it can have multiple events effect the external validity. Due to turnover and retention rates at schools for teachers, new students every year, and environmental factors/events can cause error variance to the data being collected [13]. The use of using a single year and comparing between similar schools can show the changes can be directly related to the performance evaluation model, and reduce the impact of error variance.

8.2 Participants

The participants in the study would be secondary school teachers and secondary school students. The goal of the research would be to identify a school district with multiple high schools and select two schools with a large enough population to have a sample size of 100 teachers and 1200 students. An evaluation of the cultural, racial, socio-economic, and previous academic performance would have to be gathered and make sure that school students and teachers that are part of the study are similar in demographics to try to reduce the effects of confounding variables on the correlational study. Information would be distributed to all students, parents, students, and faculty and explain that a new teacher evaluation model is being implemented at school in the coming school year, as a pilot program, to gather data, evaluate the results of the data, and then make a decision before implementing across the entire district or choosing to not change based on the gathered information.

When deciding on the school district a district will need to be selected based on data gathered from each states school report cards and boards of education. One example of school district to use would be District 212 which is made up for five high schools, Hoffman Estates High School, J B Conant High School, Palatine High School, Schaumburg High School, W M Fremd High School, which is located in Suburban Cook County, IL and has similar demographics across all five schools [2]. The research being conducted would use between subject design as the goal is to compare different groups (the different schools within the district) [24]. The experiment and data being collected would be done in a quasi-experiment design as the participants will not be random, and their participation in the study will be known as the teachers and the students are part of the school based on employment and location and will be decided by the group conducting the research and administering the new evaluation model [24].

9. Method

9.1 Measures

The two baseline measures that are highly effective to judge and perceive student learning is to evaluate reading and math abilities. To measure the effective of the training model for teachers and student growth, it would important to determine growth and over a period time and demonstrate mastery of reading and math skills along with ACT scores year over year a longitudinal improvement. Students would be evaluated at the start of freshman year, end of freshman year, then at the end of each school year for the next three years to show growth of each skill as a test-retest scenario.

One assessment tool is the ALEKS PPL (placement, preparation, and learning) which is an adaptative mathematics tool which assesses a student's readiness for math/prealgebra, beginning algebra, intermediate algebra, college algebra, precalculus, or calculus 1 [24]. The exam is 30 free response questions that are selected from a question bank that covers 314 different topics and the assessment will create a score for what a student does or does not know and what they are ready to learn next [24]. The reliability of the exam when looking at 700,000 cases found a high correlation of .96 between first scores and what would happen on a second assessment [24]. The validity of the assessment is supported by the placement of the student based on the assessment score and the class placement at level that was not too difficult or too easy [24].

A second assessment tool to evaluate reading and writing abilities would be the Oral and Written Language Scales, Second Edition: Reading Comprehension and Written [25]. The goal of this assessment is to provide a comprehensive measure of oral and written with a composite score that is designed to measure understanding of syntax, lexical/semantics, supralingusitc understanding, and pragmatics [25]. The raw scores are converted to a standard score that will create a percentile rank, age and grade scores which can also be compared to other norm referenced tests. The test is composed of 14-18 aged based items with a response time of 30 seconds per item out of a pool of 294 items to evaluated [25]. The split half reliability of the assessment is .96 and a Rasch analysis generated a reliability from .93-.99 [25]. The validity of the assessment is also very strong with a r of .86 to the WJ-III NU broad reading composite [25]. The make the application of the BARS teacher evaluation generalized to a larger population, the ACT, a national assessment, would be another assessment to evaluate the experimental group to a larger population. The ACT would be a test to show growth as the scores are deemed to have

a high level of reliability of their composite scores to GPA and college readiness of .95 for 10th graders and .96 for 11th graders [26]. The use of ACT is an indicator of college readiness and students would be able to complete college level courses [26].

9.2 Materials and Settings

The development of the BARS model for teacher evaluation would require the creation of a new teacher evaluation model that identifies behaviors a series of critical incidents by subject matter experts and each incident then scales behaviors that show a progression from lacking competency in a critical incident to mastery with clear measurable scale related to a measurable indicator and is focused on performance improvement. The goal of the BARS assessment is to design an assessment that measures critical incidents that are unique to each subject matter and identifies 5-6 critical incidents that are unique to each subject matter, along with general teaching responsibilities, and would be a good evaluation method to show progression of lacking competency to mastery of unique critical incidents. Documents would be created at each school that is a series of tables that list a critical incident with a level of proficiency on a 1-6 or 1-8 measurable scale with the lowest level that indicates no competency and the highest level which would be mastery.

9.3 Procedures: Development of BARS Evaluation Model

A BARS evaluation model is created by identifying behaviors and a series of critical incidents by subject matter experts and each incident then scales behaviors that show a progression from lacking competency in a critical incident to mastery with clear measurable scale related to a measurable indicator and is focused on performance improvement [27]. A critical incident is not an evaluation of the person, but is an observed behavior individual's on-the-job KSAO's--what happened, what action actually took place, and what were its consequences--in contrast to the various rating scales in use at the time [27]. The goal of the BARS assessment is to design an assessment that measures critical incidents that are unique to each high school subject matter. Each critical incident will be scaled on an 8-point scale with quantitative or qualitative measurements that are unique to individual subject matter and a strong BARS assessment should show examples of progression of lacking competency to mastery of unique critical incidents in your subject matter.

There are two parts to the development of the effectiveness of the model. The first step would involve assessing the students for a baseline score using the three different assessment tools to evaluate, reading, math, and college readiness. Each teacher would then need to be evaluated using the BARS teacher evaluation model to measure their individual level in each critical incident. The teacher would then meet with the rate to find areas lacking competency and areas of strength and given targets to achieve by the next observation that would move them towards mastery. The teachers would be evaluated twice per semester to evaluate growth, and an end of the school year evaluation. The goal would be to see improvement across the BARS scales to indicate higher levels of mastery of each critical incident. Students would also be assessed at the end of the school year to see if there was growth in reading, math, and college readiness. These growth scores would then be compared using a correlational study to indicate if the higher level of teacher mastery was associated with higher student performance.

9.4 Ethical, Legal, Individual, and Socio-Cultural Considerations

The development of the BARS teacher evaluation model has to follow the following criteria to meet the American Psychological Associations ethical and legal guidelines before being implemented and used. The first ethical standard is to ensure 2.01 Boundaries of Competence (a, c) as the psychologists, creation, and administration of the BARS model will have the education, training, and supervised experience, study, and professional experience along with providing the relevant education, training, supervision and consolation [28]. The development of the BARS model also involves SME's to create critical incidents and standard 2.05 Delegation of Work to Others needs to be meet when work is delegate work to employees, supervisees, or research assistants with in the schools [28]. The individuals creating the critical incidents need to be confirmed to have the competency to complete the tasks on the basis of their education, training, or experience, either independently or with the level of supervision being provided [28].

All data that is collected from both the students will need to be kept confidential to meet ethical standard 4.01 maintaining Confidentiality [28]. When also designing the new BARS teacher evaluation model ethical standard, 7.01 Design of Education and Training Programs, will ensure that the psychologists will take reasonable steps to ensure that the programs are designed to provide the appropriate knowledge and proper experiences, and to meet the requirements for licensure, certification, or other goals for which claims are made by the program [28]. The teachers that are part of the training and new BARS model will also be notified and given the purpose for the education and training programs take reasonable steps to ensure that there is a current and accurate description of the program content, training goals and objectives, stipends and benefits, and requirements that must be met for satisfactory completion of the program to meet ethical standard 7.02 Descriptions of Education and Training Programs [28].

With the creation a new assessment for teacher evaluation using the BARS model ethical standard 9.01 Bases for Assessments needs to be upheld as the opinions contained in their recommendations, reports, and diagnostic or evaluative statements need to provide information and techniques sufficient to substantiate their findings and psychologists document the efforts they made and the result of those efforts, clarify the probable impact of their limited information on the reliability and validity of their opinions, and appropriately limit the nature and extent of their conclusions or recommendations [28]. Any further consulting or supervision will

require the psychologists explain this and the sources of information on which they based their conclusions and recommendations following the completion of the BARS teacher evaluation model [28]. When the assessment model is administered ethical standard 9.02 Use of Assessments will need to be followed to ensure that the I/O Psychologists administer, adapt, score, interpret, or use assessment techniques, interviews, tests, or instruments in a manner and for purposes that are appropriate in light of the research on or evidence of the usefulness and proper application of the techniques [28]. All participants that are part of the adaption of the BARS teacher evaluation model, teacher and students who scores will be assessed are given informed consent to meet ethical standard 9.03 Informed Consent in Assessments [28, 29].

10. Discussion

The goal of all school districts, administrators, and state boards of education should be to move toward the adoption of the BARS teacher evaluation model to create quantitative teacher evaluation model that will improve professional development and lead to an increase in student performance. The new model will give clear guidance for teachers of areas of mastery and lacking competency, with clear quantitative goals/measurable bench marks will to improved teacher performance. This will lead to collection of data and the ability to find a statistical correlation between teachers' performance and student outcomes.

The adoption of the BARS model will be allow for education to adopt the principles of Industrial-Organizational psychology and use performance management strategies that are part of other industries to improve outcome and performance for all stakeholders. The development and the adoption of the new model will also provide quantitative data based upon evidence for teacher evaluations based on SME in each subject matter and will lead to greater level of reliability in evaluations, and reduce the bias from individual raters. As the measurement of teacher's performance appraisals are collected, a correlational study will be able to link specific critical incidents to outcomes of student achievement. The goal of the adoption of a BARS model will lead an increase in teacher performance, which will result in higher levels student achievement.

References

1. Bauer, T., Erdogan, B., Caughlin, D., & Truxillo, D. (2023). *Human resource management: People, data, and analytics.* Sage Publications.
2. Illinois State Board of Education. (2024). 2023 Illinois report card. *Illinois State Board of Education Publication.*
3. Birkenbach, X. C. (1984). Halo, Central Tendency, and Leniency in performance appraisel: A comparison between a graphic rating scale and a behaviourally based measure. *SA Journal of Industrial Psychology.*
4. Klieger, D. M., Kell, H. J., Rikoon, S., Burkander, K. N., Bochenek, J. L., & Shore, J. R. (2018). Development of the behaviorally anchored rating scales for the skills demonstration and progression guide. *ETS Research Report Series, 2018*(1), 1-36.
5. Conte, J. M. (2024). *Work in the 21st century: An introduction to industrial and organizational psychology*. John Wiley & Sons.
6. *Black, J. S., Bright, D. S., Gardner, D. G., Hartmann, E., Lambert, J., et al. (2019). Organizational behavior.* OpenStax
7. Reddy, L. A., Shernoff, E., & Lekwa, A. (2021). A randomized controlled trial of instructional coaching in high-poverty urban schools: Examining teacher practices and student outcomes. *Journal of School Psychology, 86,* 151-168.
8. Gupta, A., & Lee, G. L. (2020). The effects of a site-based teacher professional development program on student learning. *International Electronic Journal of Elementary Education, 12*(5).
9. Lohman, L. (2021). Evaluation of university teaching as sound performance appraisal. *Studies in Educational Evaluation, 70,* 101008.
10. Sartain, L., Stoelinga, S. R., & Brown, E. R. (2011). *Rethinking Teacher Evaluation in Chicago: Lessons Learned from Classroom Observations, Principal-Teacher Conferences, and District Implementation. Research Report.* Consortium on Chicago School Research. 1313 East 60th Street, Chicago, IL 60637.
11. Illinois State Board of Education. (2009). 2009 Illinois district report card. *Illinois State Board of Education Publication*.
12. Corcoran, S., & Goldhaber, D. (2013). Value added and its uses: Where you stand depends on where you sit. *Education Finance and Policy, 8*(3), 418-434.
13. Bordens, K. S., & Abbott, B. B. (2002). *Research design and methods: A process approach*. McGraw-Hill.
14. Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement*. Mayfield Publishing Co.
15. MBA Skool Team. (2021). Behavior Based Rating Scales, Meaning, Steps, and *MBA Skool.*
16. McCoy, J. (2023). Behaviorally Anchored Rating Scale: A Full Guide with Examples. *Academy to Innovate HR*.
17. Pazos, P., Cima, F., Kidd, J., Ringleb, S., Ayala, O., Gutierrez, K., & Kaipa, K. (2020, June). Enhancing teamwork skills through an engineering service-learning collaboration. In *2020 ASEE Virtual Annual Conference Content Access, Virtual Online.*
18. Reed, S. J. (2020). Measuring learning and engagement during debriefing: a new instrument. *Clinical Simulation in Nursing, 46*, 15-21.
19. Kogan, J. R., Dine, C. J., Conforti, L. N., & Holmboe, E. S. (2023). Can rater training improve the quality and accuracy of workplace-based assessment narrative comments and entrustment ratings? A randomized controlled trial. *Academic Medicine, 98*(2), 237-247.
20. Sjahruddin, H., Hidayat, P., Moridu, I., & Sutaguna, I. N. T. (2022). Determination of The Best Employee Using The Behavioral Anchor Rating Scale. *Jurnal Mantik, 6*(3), 2753-

2758.
21. Holland, J. R., Arnold, D. H., Hanson, H. R., Solomon, B. J., Jones, N. E., Anderson, T. W., ... & Ciener, D. A. (2022). Reliability of the Behaviorally Anchored Rating Scale (BARS) for assessing non-technical skills of medical students in simulated scenarios. *Medical Education Online, 27*(1), 2070940.
22. Urban, J. B., & van Eeden-Moorefield, B. M. (2018). *Designing and proposing your research project.* American Psychological Association.
23. American Psychological Association. (2024). Academic Writer. American Psychological Association.
24. Carlson, J. F., Geisinger, K. F., Jonson, J. L., & Anderson, N. A. (2021). The twenty-first mental measurements yearbook. (*No Title*).
25. deFur, S. & Ward, S. (2014). *Oral and written language scales (2nd Ed.). Reading comprehension and written expression*. The Nineteenth Mental Measurement Yearbook
26. Allen, J. M., & Mattern, K. (2019). Validity Considerations for 10th-Grade ACT State and District Testing. Insights in Education and Work. *ACT, Inc.*
27. Ansari, M. A., & Baumgartel, H. (1981). The critical incident techniques: Description and current uses. *Journal of Social and Economic Studies, 9*(2), 221-231.
28. American Psychological Association. (2017). Ethical Principles of Psychologists and Code of Conduct. *American Psychological Association*.
29. Debnath, D., Kell, H., Rikoon, S., Burkander, K., Bochenek, J., & Shore, J. (2018). Development of the behaviorally anchored rating scales for the skills demonstration progression guide. *Wiley*

Appendix

In the development of the BARS model I worked with various subject matter experts and helped explain the purpose behind and behavior anchored rating scale and critical incidents. The various SME created several BAR evaluations models as examples for Science, Math, Social Studies, and English

Subject Matter Expert Name: Vic Trevino
Critical Incident: Subject Matter: Math

Critical Incident: Model with Mathematics. All mathematics teachers should use examples from real world so that students learn how to APPLY math skills to solving these types real world problems. Example 1: In Geometry, a teacher can explain that a firefighter must position his ladder at the correct distance and angle to reach a two-story window. Select and apply the Pythagorean Theorem. Example 2: In Precalculus, a teacher explains that an engineer must design a new satellite dish. Select and apply the concept & formula of a parabola to the dimensional requirements for this satellite dish.

Observed Level	Explanation of Behavior/Competency
Level 1	0 – 12.5% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 2	12.5 – 25% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 3	25% – 37.5% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 4	37.5% – 50% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 5	50% – 62.5% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 6	62.% – 75% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 7	75% – 87.5% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.
Level 8	87.5% – 100% of students can correctly identify the elements in the real -world example that will direct them to the appropriate mathematical formula or skill, and then apply that skill to correctly solve the problem.

Critical Incident: Make sense of problems and PERSEVERE in solving them. PRODUCTIVE STRUGGLE CAN BE GOOD. Math teachers must teach students to persevere while problem solving. In order to do this, a teacher should provide problem sets to students that require skills that have not already been mastered and/or problems that require students to combine/synthesize previously learned skills. Example 1: In Geometry, the surface area of a rectangle is found by Base * Height. Now ask students to find the surface area of a tissue box. Example 2: Find the area left in a picture frame when the picture in the frame is not considered.

Observed Level	Explanation of Behavior/Competency
Level 1	0 – 12.5% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 2	12.5 – 25% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 3	25% – 37.5% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 4	37.5% – 50% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 5	50% – 62.5% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 6	62.% – 75% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 7	75% – 87.5% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.
Level 8	87.5% – 100% of students can correctly analyze a problem in order to know which previously learned, but not mastered, skill(s) to apply, Then apply that skill alone or in combination with other skills to solve the problem.

Critical Incident: Attend to Precision. Math teachers must require that students provide precise and exact numerical values, in each step of solving a problem and in the final answer as well. This must be done when speaking in classroom discussions and when writing, by teacher and student. The importance of this is magnified in real world situations. Example: A pharmacist must be precise when filling a prescription, because the consequences of even a small error could be severe. Example: Engineering & constructing a highway bridge, to safely hold a load of traffic

Observed Level	Explanation of Behavior/Competency
Level 1	0 – 12.5% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 2	12.5 – 25% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 3	25% – 37.5% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 4	37.5% – 50% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 5	50% – 62.5% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 6	62.% – 75% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 7	75% – 87.5% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.
Level 8	87.5% – 100% of students demonstrate mathematical precision when speaking during discussions and when writing out steps and solutions to problems.

Critical Incident: Structured Classroom Mathematical Discussion and Collabortation. Math teachers should intentionally set aside time for classroom discussion and student collaboration. Discussion and collaboration occurs often in other disciplines and it should occur regularly in math. Through open ended questions and through collaboration with teacher and classmates, students learn to how to construct viable mathematical arguments, how to brainstorm, to respectfully critique the reasoning of others, and to explain their own reasoning.

Observed Level	Explanation of Behavior/Competency
Level 1	Teacher speaks the entire time and provides no time for discussion.
Level 2	Teacher calls on the same one or two students for answers.
Level 3	Teacher asks a few random students to explain their answers.
Level 4	Teacher asks for questions from students. If there are no questions, teacher moves on to the next topic.
Level 5	Teacher asks students open ended questions based on the results of the lesson. Teacher does not move on until students respond.
Level 6	Teacher randomly assigns students to discuss a problem set, and gives them time in class to develop strategies and come up with solutions.
Level 7	Teacher intentionally groups students to discuss and collaborate on a problem set, to agree on appropriate strategies, and to report their findings back to the class.
Level 8	Student arrive to class and initiate and direct the majority of the classroom processes, discussions, and presentations to the class. Teacher participation is minimal, ideally reserved to beginning and ending class.

Critical Incident: Contextualization by citing successful Prior Learning and Deconstruction (Regularity in repeated reasoning) go Hand in Hand.

Contextualization → In math, because skills are usually stacked on successful prior learning, teachers must show students how current learning connects to prior learning and / or connects "big picture." Example: A student can use the Pythagorean theorem only after determining that a right triangle is involved. This requires prior knowledge of the definition of a right triangle. This provides "context" for the student. Why is the Pythagorean theorem so important? Because it's used in construction, architecture, navigation, surveying, etc. This is the "big picture."

Deconstruction → Teachers must show students that the regularity of repeated reasoning can direct the student to which mathematical skill to apply. Example: The same skill of working with square roots in using the Pythagorean theorem also occurs in solving circles, parabolas, trigonometric proofs, etc.

Observed Level	Explanation of Behavior/Competency
Level 1	Teacher provides no context for how current skills and concepts connect to prior learning.
Level 2	Teacher briefly mentions how current learning connects with prior learning. Teacher does not provide specific examples.
Level 3	Teacher provides one or two examples of how current mathematical skills connect to prior learning.
Level 4	Teacher provides multiple, detailed examples of how current learning and skills are connected to prior learning.
Level 5	Teacher loops back to prior topic and has students work on problems from a previous lesson that connect to current lesson problem sets, using the same mathematical skills while doing both groups.
Level 6	Teacher ask students to reflect and explain on how and in what ways current learning connects to prior learning.
Level 7	Teacher assesses students with problem sets which are only valid because they require applying identical skills as previously learned and as currently needed.
Level 8	Teacher takes the connection and context between prior learning and current learning and extends it to future topics, laying the seed and groundwork for what's to come and demonstrating "big picture" thinking.

Critical Incident: Math Teachers must show students how to reason abstractly as well as quantitatively for the express purpose of solving problems. A student must be able to analyze a problem from an abstract view, so that the student can conceptualize the scenario presented into a particular type. Once the student can characterize the scenario presented into a particular type, this will direct the student to those strategies and quantitative methods necessary to solve it. Example: Finding the amount of a discount when an item goes on sale in order to know how much it costs now. A student must analyze this scenario to know that it can involve may involve ratios, percentages, multiplication and subtraction.

Observed Level	Explanation of Behavior/Competency
Level 1	0 – 12.5% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 2	12.5 – 25% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 3	25% – 37.5% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 4	37.5% – 50% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 5	50% – 62.5% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 6	62.% – 75% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 7	75% – 87.5% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.
Level 8	87.5% – 100% of students can abstractly conceptualize a given scenario into a particular type that will direct them to the appropriate strategies and quantitative methods necessary to solve the problem presented.

Subject Matter Expert Name: Kelly Sullivan
Critical Incident: Subject Matter-Science

Critical Incident: Teacher connects science concepts to the mathematical formulas that reinforce or prove those concepts with students showing mastery by selecting correct formulas for a given set of circumstances and solves those problems correctly, representing answers with correct numerical value and units of measure.

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 2	15-30% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 3	31-45% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 4	46-55% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 5	56-65% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 6	65-75% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 7	75-87% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 8	88-100% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.

Critical Incident: Teacher instructs, models and utilizes graphing to demonstrate the relationship or lack there of between variables in an experiment or experimental design. Graphs modeled and created by teacher and students will have properly labeled x and y axis, with the independent variable being graphed on the x axis and dependent variable graphed on the y axis, the graph scaling needs to be appropriate for the data given.

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 2	16-30% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 3	31-45% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 4	46-55% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 5	56-65% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 6	66-75% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 7	76-87% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.
Level 8	88-100% of graphs created by students are properly labeled, with independent variables graphed on the x-axis and the dependent variable graphed on the y-axis, scaled appropriately.

Critical Incident: Teacher models evaluating data and the relationships between the variables measured by calculating the slope of the resulting line of the graphs, identifying that positive slope indicates a direct relationship (or positive motion), a negative slope indicates an inverse relationship (or negative motion) and no slope indicates no relationship between variables or no change in motion.

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 2	16-30% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 3	31-45% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 4	46-55% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 5	56-65% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 6	66-75% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 7	76-87% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.
Level 8	88-100% students can calculate the slope of the line of a graph and communicate verbally or in writing the direct, inverse, or lack relationship between the independent and dependent variable or describe the motion accurately.

Critical Incident: Teacher models and gives students opportunity to measure variables, create data tables that are appropriately labeled with units of measure, independent variable in column 1 and trials of data collected for the dependent variable in following columns, and indicate any equation being used to analyze data in appropriate columns

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 2	16-30% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 3	31-45% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 4	46-55% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 5	56-65% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 6	66-75% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 7	76-87% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.
Level 8	88-100% Students can create data tables that are clear and concise, correctly labeled with independent variable labeled in first column, units of measure indicated, trials and dependent variables in following columns with units of measure indicated, and any necessary equations for data analysis in finals columns.

Critical Incident: Teacher uses models and gives students opportunities to build models to show understanding of scientific concepts (build a DNA strand, create a projectile launcher, show pressure laws with aluminum cans etc) using that model to collect data when appropriate. All models should be labeled and described with images, text and calculations.

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 2	16-30% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 3	31-45% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 4	46-55% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 5	56-65% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 6	66-75% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 7	76-87% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.
Level 8	88-100% students can build a model showing the components of a structure or a model that can demonstrate a scientific principle. All models should be able to collect data when appropriate, be labeled, described with images and text, and utilize calculations when appropriate.

Subject Matter Expert Name: Robert Morowczynski
Critical Incident: Subject Matter English

Critical Incident: Dialogue in Group Exercises

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are engaged in dialogue the entire class period
Level 2	12.5% - 25% of students are engaged in dialogue the entire class period
Level 3	12.5% - 25% of students are engaged in dialogue the entire class period
Level 4	37.5%- 50% of students are engaged in dialogue the entire class period
Level 5	50%- 62.5% of students are engaged in dialogue the entire class period
Level 6	50%- 62.5% of students are engaged in dialogue the entire class period
Level 7	50%- 62.5% of students are engaged in dialogue the entire class period
Level 8	50%- 62.5% of students are engaged in dialogue the entire class period

Critical Incident: Essay Copy Editing

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are able to correctly edit over 50% of an essay
Level 2	12.5% - 25% of students are able to correctly edit over 50% of an essay
Level 3	25%- 37.5% of students are able to correctly edit over 50% of an essay
Level 4	37.5%- 50% of students are able to correctly edit over 50% of an essay
Level 5	50%- 62.5% of students are able to correctly edit over 50% of an essay
Level 6	62.5%-75% of students are able to correctly edit over 50% of an essay
Level 7	75%- 87.5% of students are able to correctly edit over 50% of an essay
Level 8	87.5%-100% of students are able to correctly edit over 50% of an essay

Critical Incident: Understanding Literary Elements

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are able to identify 5 main literary elements of a text
Level 2	12.5% - 25% of students are able to identify 5 main literary elements of a text
Level 3	25%- 37.5% of students are able to identify 5 main literary elements of a text
Level 4	37.5%- 50% of students are able to identify 5 main literary elements of a text
Level 5	50%- 62.5% of students are able to identify 5 main literary elements of a text
Level 6	62.5%-75% of students are able to identify 5 main literary elements of a text
Level 7	75%- 87.5% of students are able to identify 5 main literary elements of a text
Level 8	87.5%-100% of students are able to identify 5 main literary elements of a text

Critical Incident: Public Speaking Competence

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are able to coherently speak in front of a group
Level 2	2.5% - 25% of students are able to coherently speak in front of a group
Level 3	25%- 37.5% of students are able to coherently speak in front of a group
Level 4	37.5%- 50% of students are able to coherently speak in front of a group
Level 5	50%- 62.5% of students are able to coherently speak in front of a group
Level 6	2.5%-75% of students are able to coherently speak in front of a group
Level 7	75%- 87.5% of students are able to coherently speak in front of a group
Level 8	87.5%-100% of students are able to coherently speak in front of a group

Critical Incident: Literature Retention Accuracy

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 2	2.5% - 25% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 3	25%- 37.5% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 4	37.5%- 50% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 5	50%- 62.5% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 6	2.5%-75% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 7	75%- 87.5% of students are able to demonstrate knowledge of the text while using literary definitions and parameters
Level 8	87.5%-100% of students are able to demonstrate knowledge of the text while using literary definitions and parameters

Subject Matter Expert Name: Rich Karnia & Carl Harper
Critical Incident: Subject Matter- Social Studies

Critical Incident: Teacher connects science concepts to the mathematical formulas that reinforce or prove those concepts with students showing mastery by selecting correct formulas for a given set of circumstances and solves those problems correctly, representing answers with correct numerical value and units of measure.

Observed Level	Explanation of Behavior/Competency
Level 1	5-15% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 2	15-30% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 3	31-45% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 4	46-55% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 5	56-65% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 6	65-75% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 7	75-85% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.
Level 8	90-100% of students can correctly identify the variables given in a scientific scenario and select the proper mathematical formula necessary to solve the problem.

Critical Incident: Subject Matter Knowledge

Teacher knowledge of pedagogy is important to be able to transfer knowledge to students and help them be able to reach a higher zone of proximal development. The subject matter mastery is important for the ability to pass clear connected information to students

Observed Level	Explanation of Behavior/Competency
Level 1	Teacher has no content knowledge of the subject matter being taught
Level 2	Teacher understands content at same level as students
Level 3	Teacher understands content at same level as textbook
Level 4	Teacher understands content with a few additional ideas beyond the textbook
Level 5	Teacher understands content at above textbook level, is able to connect and relate to other ideas in the course
Level 6	Teacher understands content and is able to connect ideas to other areas of social studies, current events, and ideas beyond those covered in the course
Level 7	Teacher understands content and able to explain ideas at AP level of course depth
Level 8	Teacher has mastery of the content being taught

Critical Incident: Social Studies Exam creation.

When creating exams, it is important for the content being assess to be related to what the students have been learning, and be rooted in predetermined content/key ideas for each unit or class. The exams should have a high-level content and construct validity to be a strong measure of student master of individual subjects.

Observed Level	Explanation of Behavior/Competency
Level 1	0 % of the exam/assessment has content and construct validity
Level 2	15% of the exam/assessment has content and construct validity
Level 3	30% of the exam/assessment has content and construct validity
Level 4	45% of the exam/assessment has content and construct validity
Level 5	60% of the exam/assessment has content and construct validity
Level 6	75% of the exam/assessment has content and construct validity
Level 7	90% of the exam/assessment has content and construct validity
Level 8	90-100% of the exam/assessment has content and construct validity

Critical Incident: Student engagement in the classroom:

Teachers need to design lessons that use a variety of activities and methods to deliver instructions. The students should be engaged to the structure of a well-developed lesson.

Observed Level	Explanation of Behavior/Competency
Level 1	0%-12.5% of students are engaged the entire class period
Level 2	12.5% - 25% of students are engaged the entire class period
Level 3	25%- 37.5% of students are engaged the entire class period
Level 4	37.5%- 50% of students are engaged the entire class period
Level 5	50%- 62.5% of students are engaged the entire class period
Level 6	62.5%-75% of students are engaged the entire class period
Level 7	75%- 87.5% of students are engaged the entire class period
Level 8	87.5%-100% of students are engaged the entire class period

Critical Incident: Presentation of Material/Lecture

A key component of a social classroom is being able to develop and deliver a wide amount of material in a meaningful and engaging method to students. Presentations should be well organized, and clear to follow big ideas of major concepts/content.

Observed Level	Explanation of Behavior/Competency
Level 1	No organized lesson or prestation is present in observation
Level 2	Prestation is very poorly organized, lacks structure, and is difficult to follow
Level 3	Presentation is simple, lists major concepts and key ideas, but lacks any explanations of ideas
Level 4	Presentation is organized, has structure, but is filled with too much information per slide to be understood
Level 5	Presentation is simple, lists major concepts and key ideas, and provides concise explanation of key ideas, but lacks any additional support or visual aides
Level 6	Presentation is simple, lists major concepts and key ideas, and provides concise explanation of key ideas, has additional support and visual aides to enhance learning
Level 7	Presentation is well developed, lists major concepts and key ideas, and provides concise explanation of key ideas, has additional support or visual aids, but provides no additional interaction for students
Level 8	Presentation is well developed, lists major concepts and key ideas, and provides concise explanation of key ideas, but lacks any additional support or visual aids, and serves as an interactive learning tool to enhance student engagement