Title: Towards Measures of Different and Useful Aspects of Schooling: Why Schools Need Both Teacher Assigned Grades and Standardized Assessments

Abstract:
Teacher assigned grades are well-known to be highly predictive of high school graduation, college enrollment and college completion, but there has been little research as to why. This study examines the extent to which grades in high school are a systematic measure of student effort, participation and behavior that is a different and useful measure for schools and school leadership beyond standardized test scores. Using hierarchical linear modeling to analyze the ELS:2002 dataset, this study demonstrates that a significant percentage of the variance in grades is between schools, with teacher ratings of student engagement accounting for a large proportion of grades, which has important implications for the policy and practice of standardized assessment and grade use in schools.

Purpose and Background:
The purpose of this study is to examine the extent to which teacher assigned grades in high school represent a useful measure of schooling, above and beyond standardized test scores. Historically, grades have been maligned by pyschometricians for their "hodgepodge" nature, in which when asked what they assign a grade for, teachers respond that grades are assigned for a multitude of outcomes, such as academic knowledge, student participation, effort, and behavior (Cross & Frary, 1999), known as "kitchen-sink" grading (Cizek, Fitzgerald, & Rachor, 1995-1996). Some scholars have interpreted this to mean that grades are subjective and unreliable measures of academic performance, and thus must be reformed to align much more to standardized test scores (Brookhart, 1991, 2011). As recently noted, "student's grades often have little relation to their performance on state assessments (p.23)" (Guskey & Jung, 2012). But *should* grades have a relation to standardized test performance? If test scores are assumed to be an accurate and reliable measure of fundamental academic knowledge, why would schools need another measure of this factor? The purpose of schooling in the U.S. is far from agreed upon (Labaree, 1997) and some have argued that test scores are a poor measure of what the many different stakeholders in schools are looking for schools to instill in their students (Nichols & Berliner, 2007). Might it be that grades measure different, but important aspects of schooling?

Standardized test scores have historically lacked criterion validity to overall schooling outcomes (Atkinson & Geiser, 2009), to such an extent that many states throughout the U.S. have begun to mandate exit and end of course exams (Allensworth, 2005a; Blazer, 2012; Nichols & Berliner, 2007; Warren, Jenkins, & Kulick, 2006) which artificially connect test scores to outcomes through retention, grade promotion and graduation requirements. By contrast, teacher assigned grades are strong predictors of overall schooling outcomes, such as graduation or dropping out (Allensworth, 2005b; Barrington & Hendricks, 1989; Battin-Pearson et al., 2000; Bowers, 2010; Bowers & Sprott, 2012; Bowers, Sprott, & Taff, 2013; Lloyd, 1978) as well as college attendance and graduation (Atkinson & Geiser, 2009; Cliffordson, 2008). In addition, grades are seen as being "fairer" assessments than standardized tests, since grades are not as strongly related to socio-economic status (SES) than test scores (Atkinson & Geiser, 2009). As noted by Atkinson & Geiser (2009) "High school grades are sometimes viewed as a less reliable indicator than standardized tests because grading standards differ across schools. Yet although grading

standards do vary by school, grades still outperform standardized tests in predicting college outcomes." (p.665)

The focus of the present study is to thus ask the question: why? What is it about grades that make them a strong predictor of overall schooling outcomes that adds to the knowledge gained about student learning from standardized test scores? If schools have two measures of different and useful factors about different student outcomes from schooling, then schools should use both sets of measures to inform their practice and decision making (Bowers, 2009, 2011; Farr, 2000).

An emerging research domain has begun to examine this issue of the relationship between teacher assigned grades and standardized assessment scores. Over the past 30 years, numerous studies have demonstrated that across multiple contexts, as well as nationally, grades and standardized test scores correlate at about 0.5 (Bowers, 2011; Brennan, Kim, Wenz-Gross, & Siperstein, 2001; Duckworth, Quinn, & Tsukayama, 2012; Linn, 1982, 2000). Thus, when squaring this correlation, the R-squared indicates that about 25% of what grades are is the fundamental academic knowledge assessed by standardized test scores (Bowers, 2011). But then what is the other 75% of grades, and how does the majority of the variance in grades relate so well to overall schooling outcomes? A growing set of research studies over the past decade has postulated that grades are multidimensional (Bowers, 2011), assessing academic knowledge to a limited extent, but more importantly, assessing what has been termed a "conative" factor (Willingham, Pollack, & Lewis, 2002), a "common grade dimension" (Klapp Lekholm, 2011; Klapp Lekholm & Cliffordson, 2008, 2009; Thorsen & Cliffordson, 2012), and a "Success at School Factor (SSF)" (Bowers, 2009, 2011). Across these studies, other than academic knowledge, grades appear to measure student engagement through measuring effort, participation and behavior. Recent research has confirmed that that while grades reflect student self-perception, self-efficacy, and self-control across subjects (Klapp Lekholm & Cliffordson, 2009), these factors are mediated through teacher evaluations of student conduct and homework completion (Duckworth, et al., 2012). Thus, these findings indicate that beyond the assessment of the academic knowledge reflected in standardized test scores, what teachers assess with grades is student engagement, effort, participation and behavior, which reflect measures of student self-control and self-efficacy. This research postulates that it is these factors that give grades their predictive validity with overall schooling outcomes, since if grades are a valid measure of how well a student can negotiate the non-academic components of the schooling process, then it is these factors that predict later student ability to conform to the institutional expectations that lead to completing high school as well as post-secondary schooling and employment (Bowers, 2011). However, to date these studies have been mostly limited to either examining the entire population of students in Sweden (Klapp Lekholm, 2011; Thorsen & Cliffordson, 2012), or to small intact samples.

Framework of the Study:

Thus, the central aim of the present study is to examine the extent to which teacher assigned grades are a useful assessment of student engagement, using a large nationally generalizable sample of U.S. grade 10 high school students. This study will examine three main aspects of this issue. First, to date, while the standardized grading practices literature claims that grades are unreliable and subjective measures that vary too much across schools to be useful, very little research has been done to examine the extent to which grades actually do vary within and

between schools. Second, while critics of standardized assessments note that socio-economic status and ethnicity are strongly associated with test scores, little work has been done to examine the extent to which grades, test scores and SES are related, and to what extent grades may be a fairer, or more "just" assessment that does not vary as strongly by SES or the demographic background of the student as do standardized assessments. Third, once these two main issues with control variables are addressed (within/between school variance and student SES/background variables) the remaining variance in grades that is not explained by standardized test scores can be examined as to the extent that teacher evaluation of student effort, participation and behavior are associated with the grades they assign, and is this assessment consistent across schools, and thus more reliable than previously inferred from the past standardized grading practices literature?

Research questions:

1. To what extent do grade 10 report card grades vary between and within schools in the U.S.?
2. To what extent are grade 10 report card grades related to standardized test scores and student socio-economic status?
3. To what extent are teacher perceptions of engagement, through effort, participation and behavior, associated with grade 10 report card grades?

Methods:

Data:

This study is a secondary analysis of the restricted use Education Longitudinal Study of 2002 (ELS:2002) dataset. ELS:2002 was originally collected by the National Center for Education Statistics (NCES), in which about 15,400 U.S. grade 10 students across 750 school in 2002 were surveyed on a large array of items concerning their high school experience, as well as collecting demographic information, standardized assessments in mathematics and reading that were aligned to NAEP and PISA, and students report card grades and overall GPA (Ingles et al., 2007). In addition, NCES surveyed the student's English and mathematics teachers from the 2001/2002 academic year asking the teachers about each student's performance in their courses. As will be further detailed in the final paper, and as provided in the proposal appendix in Table 1, I included the non-cumulative grade point average across all courses for students in grade 10 as well as grade 10 mathematics and reading standardized tests scores and a range of student and school background variables as well as teacher ratings of student engagement. In addition, because ELS is not a simple random sample, but is a probabilistic complex sample, I applied the sampling weights to allow for generalization to all three million students who were in grade to in the U.S. in 2002.

Analysis:

To address the research questions I used hierarchical linear modeling (HLM) (Hox, 2010; Raudenbush & Bryk, 2002) in SPSS (Heck, Thomas, & Tabata, 2010) to examine two models with fixed effects. For both HLM analyses, the dependent variable is non-cumulative grade 10 GPA, which is the average of a student's grades across all subjects from only grade 10. Preliminary factor and reliability analysis indicated that neither the English nor the mathematics teacher variables loaded onto similar factors. Thus, I included each variable individually.

Results:

As will be further explicated in the full final paper, there were three main findings. First, the unconditional HLM indicated that there is a significant amount of variance in grades between schools (Wald Z = 13.390, $p<0.001$) with the interclass correlation coefficient (ICC) suggesting that 16.52% of the variance in grade 10 GPA is between schools. Second, in the Appendix, Table 2 presents the results of the two HLM analyses. In Model A, only student and school-level background and context variables are included, which account for 36.83% of the variance at the student level and 45.54% of the variance at the school level. In Model B, English and mathematics teacher ratings of student effort, participation and behavior explained an additional 33.17% of the variance in grade 10 GPA at the student level and an additional 13.49% at the school level (subtract Model B variance explained from Model A at each level). These results indicate that controlling for test scores and background and demographic variables at the student and school level, teacher evaluations of student effort, participation and behavior make up a significant portion of what grades represent in grade 10, beyond test scores. Third, in examining the individual parameter estimates in the full final Model B in Table 2, the only significant ethnicity variable is Native American, and the standardized coefficient (beta) for SES is relatively small, which is in stark contrast to the literature on these variables as they relate to standardized test scores. Third, the estimates of multiple other variables are of interest and will be discussed at length in the final paper. As an example, in replication of the Swedish studies (Thorsen & Cliffordson, 2012), females received higher grades on average than males (0.108 grade points) controlling for the other variables in the model.

Significance & Conclusion:

This study addresses the conference theme of *The Power of Education Research for Innovation in Practice and Policy* through providing a novel analysis of the extent to which teacher assigned grades vary within and between schools in the U.S., the relationship of SES and demographics to grades and tests, and the extent that teacher evaluations of student effort, participation and behavior are significant and systematic components of grades. Using a large nationally generalizable sample of high school students, this study demonstrates that grades are a useful and consistent measure across schools. Moreover, the significant portion of variance in grades that lies between schools indicates that grades have a systematic component, and that when compared with past models of the relationship between SES and test scores, grades appear to be somewhat less influenced by SES. Because the models control for test scores, SES and demographic variables, Model B examines for the first time in an HLM framework the extent that teacher ratings of effort, participation and behavior are associated with the previously postulated conative/SSF/common grades dimension suggested by the previous literature, and that these teacher perceptions account for a large proportion of this component of grades.

References:

Allensworth, E. M. (2005a). Dropout rates after high-stakes testing in elementary school: A study of the contradictory effects of Chicago's efforts to end social promotion. *Education Evaluation and Policy Analysis, 27*(4), 341-364. doi: 10.3102/01623737027004341

Allensworth, E. M. (2005b). Graduation and dropout trends in Chicago: A look at cohorts of students from 1991 through 2004 Retrieved July 7, 2006, from www.consortium-chicago.org/publications/p75.html

Atkinson, R. C., & Geiser, S. (2009). Reflections on a Century of College Admissions Tests. *Educational Researcher, 38*(9), 665-676. doi: 10.3102/0013189x09351981

Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research, 82*(6), 309-319.

Battin-Pearson, S., Abbott, R. D., Hill, K. G., Catalano, R. F., Hawkins, J. D., & Newcomb, M. D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92*(3), 568-582. doi: 10.1037/0022-0663.92.3.568

Blazer, C. (2012). National trends in end-of-course assessment programs: Miami-Dade County Public Schools.

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration, 47*(5), 609-629. doi: 10.1108/09578230910981080

Bowers, A. J. (2010). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *Journal of Educational Research, 103*(3), 191-207. doi: 10.1080/00220670903382970

Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation, 17*(3), 141-159. doi: 10.1080/13803611.2011.597112

Bowers, A. J., & Sprott, R. (2012). Examining the Multiple Trajectories Associated with Dropping Out of High School: A Growth Mixture Model Analysis. *Journal of Educational Research, 105*(3), 176-195. doi: 10.1080/00220671.2011.552075

Bowers, A. J., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal, 96*(2), 77-100.

Brennan, R. T., Kim, J., Wenz-Gross, M., & Siperstein, G. N. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review, 71*(2), 173-215.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 10*(1), 35-36.

Brookhart, S. M. (2011). *Grading and Learing: Practices that support student achievement*. Bloomington, IN: Solution Tree Press.

Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995-1996). Teachers' assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment, 3*(2), 159-179.

Cliffordson, C. (2008). Differential Prediction of Study Success Across Academic Programs in the Swedish Context: The Validity of Grades and Tests as Selection Instruments for Higher Education. *Educational Assessment, 13*(1), 56-75. doi: 10.1080/10627190801968240

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What *No Child Left Behind* leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology, 104*(2), 439-451. doi: 10.1037/a0026280

Farr, B. P. (2000). Grading practices: An overview of the issues. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 1-22). Norwood: Christopher-Gordon Publishers.

Guskey, T. R., & Jung, L. A. (2012). Four steps in grading reform. *Principal Leadership, 13*(4), 22-28.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and Longitudinal Modeling with IBM SPSS*. New York, NY: Routledge.

Hox, J. (2010). *Multilevel analysis: Techniques and applications* (Second ed.). New York: Routledge.

Ingles, S. J., Pratt, D. J., Wilson, D., Burns, L. J., Currivan, D., Rogers, J. E., & Hubbard-Bednasz, S. (2007). *Education longitudinal study of 2002: Base-year to second follow-up data file documentation*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

Klapp Lekholm, A. (2011). Effects of School Characteristics on Grades in Compulsory School. *Scandinavian Journal of Educational Research, 55*(6), 587-608. doi: 10.1080/00313831.2011.555923

Klapp Lekholm, A., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation, 14*(2), 181-199.

Klapp Lekholm, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation, 15*(1), 1-23. doi: 10.1080/13803610802470425

Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal, 34*(1), 39-81. doi: 10.3102/00028312034001039

Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies* (pp. 335-388). Washington DC: National Academy Press.

Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.

Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement, 38*(4), 1193-1200.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral Damage: How High Stakes Testing Corrupts America's Schools*. Cambridge: Harvard Education Press.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks: Sage.

Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation, 18*(2), 153-172. doi: 10.1080/13803611.2012.659929

Warren, J. R., Jenkins, K. N., & Kulick, R. B. (2006). High School Exit Examinations and State-Level Completion and GED Rates, 1975 Through 2002. *Educational Evaluation and Policy Analysis, 28*(2), 131-152. doi: 10.3102/01623737028002131

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement, 39*(1), 1-37.

Table 1: Descriptive statistics

	Mean	(SD)	Min	Max	ELS:2002 variable label and description
GPA for all 10th grade courses	2.67	0.87	0	4	F1GPA10: Non-cumulative grade 10 GPA all courses
Grade 10 Mathematics	50.71	9.91	19.38	86.68	BYTXMSTD: Grade 10 mathematics stand. T-score
Grade 10 Reading	50.53	9.89	22.57	78.76	BYTXRSTD: Grade 10 reading stand. T-score
SES	0.03	0.74	-2.12	1.87	F1SESR: Student socio-economic status
Female	0.50	0.50	0	1	BYSEX = 1 (male ref. group)
African American	0.17	0.38	0	1	BYRACE2 = 1
Student is Hispanic	0.15	0.35	0	1	BYS15 = 1
Asian	0.13	0.33	0	1	BYRACE3 = 1
Hawaiian/Pacific Islander	0.02	0.14	0	1	BYRACE4 = 1
Native American	0.04	0.21	0	1	BYRACE5 = 1
English is native language	0.83	0.38	0	1	BYSTLANG = 1
Non-Traditional family	0.41	0.49	0	1	BYFCOMP > 1: Both birth parents not present in home
English Teacher rating					
Student works hard for good grades	0.69	0.46	0	1	BYTE04: 0=no, 1=yes
How often student completes homework	3.01	1.01	0	4	BYTE13: 0=never, 1=rarely, 2=some of the time, 3=most of the time, 4=all of the time
How often student is absent	1.16	0.72	0	4	BYTE14: (same as previous)
How often student is tardy	0.63	0.84	0	4	BYTE15: (same as previous)
How often student is attentive in class	2.95	0.88	0	4	BYTE16: (same as previous)
How often student is disruptive in class	0.59	0.87	0	4	BYTE17: (same as previous)
Mathematics Teacher rating					
Student works hard for good grades	0.68	0.47	0	1	BYTM04: 0=no, 1=yes
How often student completes homework	2.99	1.02	0	4	BYTM13: 0=never, 1=rarely, 2=some of the time, 3=most of the time, 4=all of the time
How often student is absent	1.15	0.70	0	4	BYTM14: (same as previous)
How often student is tardy	0.58	0.80	0	4	BYTM15: (same as previous)
How often student is attentive in class	2.96	0.89	0	4	BYTM16: (same as previous)
How often student is disruptive in class	0.55	0.84	0	4	BYTM17: (same as previous)
School-level variables					
Urban	0.34	0.47	0	1	URBAN = 1 (rural ref. group)
Suburban	0.34	0.47	0	1	URBAN = 2 (rural ref. group)
% Free Lunch	24.51	19.13	0	96.2	CP02PLUN
% Minority students	34.36	31.20	0	100	CP02PMIN
Student teacher ratio	16.62	4.25	4.39	40	CP02STRO
Enrollment (in thousands)	1.27	0.84	0.02	4.64	CP02STEN/1000

Table 2: Hierarchical linear models examining grade 10 GPA

	Model A			Model B		
Parameter	Coeff.	β	SE	Coeff.	β	SE
Student-level variables						
Grade 10 Mathematics	0.032***	0.371	0.001	0.021***	0.235	0.001
Grade 10 Reading	0.015***	0.168	0.001	0.009***	0.103	0.001
SES	0.166***	0.142	0.011	0.085***	0.073	0.010
Female	0.303***	0.175	0.013	0.108***	0.062	0.012
African American	-0.066**	-0.029	0.021	-0.013		0.020
Hispanic	-0.019		0.027	0.039		0.025
Asian	0.088*	0.034	0.034	0.054		0.032
Hawaiian/Pacific Islander	-0.062		0.054	-0.054		0.057
Native American	-0.092**	-0.022	0.030	-0.064*	-0.015	0.027
English is native language	-0.147***	-0.064	0.026	-0.015		0.025
Non-Traditional family	-0.133***	-0.076	0.014	-0.054***	-0.031	0.012
English Teacher rating						
Student works hard for good grades				0.208***	0.111	0.018
How often student completes homework				0.153***	0.179	0.009
How often student is absent				-0.088***	-0.074	0.010
How often student is tardy				0.008		0.009
How often student is attentive in class				0.055***	0.055	0.010
How often student is disruptive in class				-0.008		0.008
Mathematics Teacher rating						
Student works hard for good grades				0.163***	0.088	0.018
How often student completes homework				0.144***	0.169	0.009
How often student is absent				-0.077***	-0.062	0.010
How often student is tardy				-0.028**	-0.025	0.009
How often student is attentive in class				0.064***	0.066	0.010
How often student is disruptive in class				0.030**	0.029	0.008
School-level variables						
Urban	-0.076		0.046	-0.051		0.042
Suburban	-0.023		0.036	-0.006		0.032
% Free lunch	0.004**	0.086	0.001	0.004***	0.096	0.001
% Minority students	-0.002*	-0.062	0.001	-0.002**	-0.084	0.001
Student Teacher ratio	0.006		0.004	0.011**	0.053	0.004
Enrollment in thousands	-0.099***	-0.096	0.022	-0.081***	-0.078	0.021
Intercept	0.325		0.083	-0.230**		0.087
Percentage of variance explained						
at student level	36.83			70.00		
at school level	45.54			59.03		
BIC	22000.13			9211.69		

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION

Reviews

Summary of reviewers notes and ratings on criteria.

Toward Measures of Different and Useful Aspects of Schooling: Why Schools Need Both Teacher-Assigned Grades and Standardized Assessments

Unit / Sub Unit: **Division L - Educational Policy and Politics / Section 3: Curriculum, Testing, and Instructional Practice**

Review Worksheet **Paper Submission: Review Panel** **(3/3)**

Review #801228

Criteria	Rate
Objectives or purposes	4 / 5
Perspective(s) or theoretical framework	5 / 5
Methods, techniques, or modes of inquiry	4 / 5
Data sources, evidence, objects or materials	5 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	4 / 5
Scientific or scholarly significance of the study or work	3 / 5

Comments to the Author/Submitter

This is an interesting paper on student grades. The author motivates the paper well and the analysis seems to be carefully done. I had three questions/comments: - The author asserts that grades are a stronger predictor of ultimate outcomes than standardized test scores, and cites important studies on this. Is it possible to look at some of these questions using the dataset at hand. For example, the author explains that test scores and things like effort both predict variation in grades. If we look at college-going or college graduation, for example, do grades predict these outcomes over and above both test scores and these other measures (teacher reports of effort, etc.)? It would be interesting to see whether there are other important dimensions of grades not captured by these other measures. - Rather than relying on the literature about racial and SES differences in standardized test scores (bottom of the Results section), it would be useful to present these results in the sample. - A minor point: I worry about the author's characterization at the top of page 3 that grades may be fairer or more just simply because they are less correlated with SES/demographics. The question is the underlying true distribution of the latent construct the grades/tests are measuring, which may well vary by SES/demographics.

Review #801233

Criteria	Rate
Objectives or purposes	5 / 5

Perspective(s) or theoretical framework	2 / 5
Methods, techniques, or modes of inquiry	3 / 5
Data sources, evidence, objects or materials	4 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	2 / 5
Scientific or scholarly significance of the study or work	5 / 5

Comments to the Author/Submitter

I didn't find the motivation behind this paper compelling. For example, the author writes "If test scores are assumed to be an accurate and reliable measure of fundamental academic knowledge, why would schools need another measure of this factor?" Well, one possible answer is that grades provide a feedback loop for students so that they know which areas they are struggling with so they can study for a state assessment. Nevertheless, I find the research area around grades and test scores to be compelling, so I think the theory and motivation to be reworked.

Review #801229

Criteria	Rate
Objectives or purposes	4 / 5
Perspective(s) or theoretical framework	4 / 5
Methods, techniques, or modes of inquiry	4 / 5
Data sources, evidence, objects or materials	5 / 5
Results and/or substantiated conclusions or warrants for arguments/point of view	3 / 5
Scientific or scholarly significance of the study or work	4 / 5

Comments to the Author/Submitter

this paper is very interesting and promising.