Understanding Secondary Teachers' Formative Assessment Practices and
Their Relationship to Student Motivation

James H. McMillan

Jessye Cohen

Lisa Abrams

Kathleen Cauley

Gina Pannozzo

Jessica Hearn

Virginia Commonwealth University

January 5, 2010

While it is clear that formative classroom assessment is receiving more attention as a critical component of teaching that directly affects student learning (Brookhart, 1997, 2004, 2007; McMillan, 2007; Shepard, 2005), there is still a lack of systematic research that addresses the nature of formative assessment and the relationship between formative assessment and student motivation. There is a lack of agreement on the nature of formative assessment (Cizek, 2010). Some view formative assessment as testing and giving feedback, similar to materials many large testing companies purport to provide teachers and schools, while others claim that it is integrated as instruction proceeds and only formative if subsequent instruction is provided (Wiliam, 2010). According to Wiliam and Leahy (2007), formative assessment "needs to contain an implicit or explicit recipe for future action." (p.13). Assessments that are formative include appropriate identification of student weaknesses or lack of understanding, contain a clear indication of the criteria used to judge performance (the ultimate learning target), measure successful progress, provide specific feedback, and include instructional correctives, different from previous instruction, that will move the student toward successful attainment of learning targets (Brookhart, 2007; Guskey, 2010). In other words, assessment that is formative provides individualized feedback and information about what students need to do to improve performance. However, there is little research on the extent to which these aspects of formative assessment are used by secondary teachers. Brookhart and Durkin (2003) researched secondary teachers' assessment practices and student motivation, but did not examine different aspects of formative assessment.

Student motivation, a critical intervening variable between formative assessment and learning, consists primarily of student engagement or effort, goal orientation and self-

efficacy. Research has shown that the goal theory perspective (Ames, 1992; Anderman and Maehr, 1994) may explain relationships between assessment and motivation. There is conclusive evidence that certain instructional practices and strategies can enhance a mastery or learning orientation in the classroom (Ames, 1992; Brookhart & Durkin (2005), and there is little reason to conclude that classroom assessments, used formatively, would not also enhance a mastery orientation. In contrast, students with a performance orientation are motivated to achieve so that positive outcomes, such as grades and test scores, are high, with avoidance of failure. This suggests that formative assessment needs to be conducted in a manner that focuses on mastery and not performance goals.

Self-efficacy refers to the student's judgment of his or her capability to organize and execute a plan of action to attain a goal. Generally, self efficacy is sensitive to variation in the performance context and the mastery criterion of performance (Zimmerman, 2000), and as such, is presumed to play a causal role in motivation. Generally, the literature suggests that the stronger a student's self-efficacy, the stronger his or her level of motivation, effort, and perseverance (Pintrich & Schunk, 2002). Finally, Shepard (2000) summarized research that showed students who practice self-evaluation are more motivated and interested in substantive feedback than students who do not self-evaluate. As students internalize the criteria for evaluating their work they are better able to connect their performance with their preparation, which enables the development of an internally oriented, controllable sense of self-efficacy (Andrade, 2010; Stiggins, 2005).

The purpose of the current study was to investigate different aspects of formative assessment, to determine teachers' self-reported formative assessment practices, and relate these practices to student self-reported motivation. Specific research questions included the following:

1. What is the nature of secondary teachers' formative assessment practice?

2. To what extent do secondary teachers use different aspects of formative assessment?

3. What differences exist in teachers' formative assessment practices according to gender, grade level, subject matter, and student ability level in the class?

4. What are the relationships between secondary teachers' use of formative assessment and student motivation?

Methodology

Sampling

In the fall of 2006, superintendents of four Richmond Virginia area school districts agreed to participate in the study. In order to represent a diverse sample of students and teachers, schools were purposively identified and contacted by researchers. The researchers sought a diverse sample that would reflect various years of teaching experience, different levels of student ability, subjects, and grade levels. Individual schools elected to participate in the study, resulting in a volunteer sample of schools from a larger number initially identified by each superintendent.

Participants

The 3,242 sixth through twelfth-grade student participants in this study came from three middle schools (*n*= 1,764), and two high schools (*n* = 1,478). The students were

evenly distributed across grade levels, with the eighth grade student sample being the largest (16.6%), and the eleventh grade representing the smallest part of the sample (7%).

The student sample was fairly diverse – with 52% of students reporting their ethnicity as White ($n = 2,113$), 34% African-American ($n = 1,367$), 5% Hispanic ($n = 206$), 3% Asian-American ($n = 115$), and 1% reporting their ethnicity as Native-American ($n= 55$). The remaining 5% of students indicated that the "other" category best described their ethnicity ($n = 195$). The participants were evenly distributed between males and females.

There were a total of 161 participating secondary teachers including 75 middle and 86 high school teachers. Teachers reported their ethnicity as White ($n = 183$, 88%), African-American ($n = 21$, 10%), and Hispanic ($n = 2$, 1%). The remaining teacher selected "Other." The majority of teachers had more than 23 years of classroom experience with the average for the sample being 14 years. The majority taught standard or regular classes (55%) and 27 percent reported teaching advanced or honors courses. Teachers working with basic skills/remedial students comprised eight percent and those in blended/inclusion settings accounted for ten percent. The majority of participants were female (80%).

Measures

To understand the relationship between teachers' grading practices and student variables, two self-report surveys were used. The student survey was used to determine levels of student motivation, engagement, self-efficacy, and goal orientation. It also evaluated student perceptions of teacher's grading and assessment practices. There were 28 items, using a five-point Likert scale, ranging from 1 (strongly disagree) to 5 (strongly

4

agree). The 28 items were adapted from existing measures. Initial survey items for engagement were adapted from the Student Participation Questionnaire (SPQ) (Finn, Folger, & Cox, 1991). The Patterns of Adaptive Learning Survey (PALS) (Midgley, Maehr, Hruda, Freeman, Gheen, Kaplan, Kumar, Middleton, Nelson, Roeser, & Urdan, 2000) was adapted slightly to measure student mastery or performance goal orientations, and self-efficacy. Student engagement was measured using ten items adapted from the 28 item Student Participation Questionnaire (SPQ) (Finn, Folger & Cox, 1992). The items combine effort, initiative, and non-participatory behavior to reflect student engagement.

To examine the underlying factor structure of the student survey, scores on the 28 items of the student survey were analyzed using principal components extraction with orthogonal rotation (varimax) to confirm the presence of the aforementioned subscales. The initial extraction revealed six components with eigenvalues over 1.0. Items were retained for later analyses if they had a factor loading of .50 or higher and if the loading on another factor was no higher than .30, or if they were theoretically relevant. This resulted in 22 items for a second EFA. The final scale included four components, which accounted for 54% of the variance. Items and factor loadings are reported in Table 1. Subscale reliabilities are presented in Table 2.

Also, students reported on their perceptions of teachers' formative assessment practices. Four items measured teacher practices concerning student self-assessment, teacher feedback and praise, and teachers' written comments.

The teacher survey included 60 five-point Likert scale items that measured what information teachers used to determine student grades (i.e., homework, class

participation, the degree to which students pay attention in class, etc and to what extent teachers), and to what extent this information was used. Teachers also indicated their use of various assessment and grading practices, (e.g., feedback, checklists to evaluate student work, and retakes of tests).   Teachers responded based upon their level of agreement with instructional and assessment practices. Examples include "I give special privileges to students who do the best work in the class" and "I tell students it's important to learn what is being taught because understanding the material is important." The teacher survey was adapted from a prior study on grading and assessment practices (McMillan, 2001; 2002).   Eight demographic questions provided additional information.

Seven of the 60 items were used to assess teachers' use of formative assessment (see Table 3). The results were analyzed for each of the items as well as for an overall formative score of the sum of the seven items. Cronbach's alpha of the seven items as a scale was .78.

Procedures

During the fall of 2006, local school districts were contacted to determine their level of interest in the study. Researchers arranged to meet with teachers during their monthly staff meeting to explain the purpose of the research, the administration procedures to answer questions, and to assure confidentiality and voluntary participation. Teachers received a packet of information containing a teacher letter, teacher consent form, instructions for administration, parent letters, student assent forms, and surveys. Teachers sent home the parent information letters and obtained student assent about two weeks prior to the survey administration. Teachers administered the student surveys in their second period class. Each survey was coded in a way that would preserve student

6

and teacher anonymity, while enabling researchers to link the student and teacher survey data. Each teacher's self-report survey was collected in a separate envelope. Researchers returned to the schools to pick up all the completed surveys.

A key aspect of the procedure was that **s**tudents were asked to complete the surveys in their classroom for only that specific class that they were taking that semester, and teachers also responded with that class in mind. Thus, the responses of the students toward the class the teachers taught could be captured, assuming that students did in fact focus their answers on what was happening in that class and not more generally on that teacher or on all their classes. Class averages were computed for students so that the unit of analysis is the class, not students, in relating student motivation to teacher formative assessment practices. While there was no indication that students were not limiting their reference to a single class, there was also no independent confirmation that this was the case.

Results

Descriptive statistics for the teacher survey responses for seven formative assessment items are summarized in Table 3. The results indicate that while about 70% of the teachers indicated that they used some kind of formative assessment, mostly in-class monitoring of student progress, formative assessment generally, and specific, individualized feedback, lower percentages are indicated for other specific formative assessment practices. Only 23% and 32%, respectively, indicated that they use assessments to diagnose student weaknesses and to guide further instruction "quite a bit" or "extensively." Half or more of the teachers reported using four important formative assessment practices "not at," "minimally," or "some," including assessments that

diagnosed student weaknesses, assessments used to guide further instruction, feedback that contained suggestions for further learning, and feedback that was given privately to each student. The standard deviation of responses for the five-point Likert scale is around 1.00, which would be expected, with some skewness to the left for means 3.70 or above, and skewness to the right for means below 3.00. There were no statistically different means when comparing teachers by gender, subject taught, grade level, or student ability.

Insert Table 3

Pearson product moment correlations were used to show the relationship between students' motivation as reported in the four scales and teachers' formative assessment practices (Table 4). Examining results for individual items revealed positive, nonsignificant correlations between students' overall goal orientation and teachers' formative assessment practices. Student self-efficacy showed a statistically significant relationship ($p < .05$) with formative assessment and assessment to diagnose student weaknesses ($r = .17$). The only other single-item correlation that was significant showed a positive relationship between the mastery scale and assessment to diagnose student weaknesses. The positive correlations between an overall formative assessment scale score and the motivation subscales were all statistically significant, though small.

Insert Table 4

Relationships were also examined between items measuring students' perceptions of teacher practices and student motivation subscales. As summarized in Table 5, most of these positive correlations were statistically significant, some showing moderate relationships. The item most related to formative assessment, "Teacher's written

comments help me understand," was significantly correlated to all four scales, notably to the mastery scale ($r =.48$) and the efficacy subscale ($r =.34$). Several moderate positive correlations were found for two items stressing the "importance" of learning and student motivation, especially for the importance of learning to understand and mastery ($r = .52$).

Insert Table 5

Discussion

The descriptive results of the reported teacher use of formative practices illustrates that many teachers in this study make limited use of aspects of formative assessment practices, especially those related to using assessment to guide further instruction, diagnose student weaknesses, and give feedback that contains suggestions to students for further learning. In light of the fact that 60% of the teachers reported extensive or moderate use of formative assessment in general, there may be a disconnect between what teachers believe constitutes formative assessment and formative assessment practices supported in the literature. At the very least, teachers appear to give insufficient attention to the use of assessment to guide instruction and learning, not simply use information to gauge current student understanding.

It was interesting to find no differences in formative assessment practices when the results were analyzed for differences between grade level, subject, student ability, and gender. This suggests that formative assessment practices are not aligned with these factors, that formative assessment is understood by teachers as a general construct; There may be little differentiation of formative assessment practices by grade level or subject, though the sample from this study is small to avoid a Type II error. Finally, the descriptive results show that there is much work needed to enhance teachers' use of formative assessment practices. One would hope that very high percentages of teachers

would use at least some of the techniques, such as giving individual feedback and ongoing monitoring of student progress.

Given all the influences that effect student motivation, much of which are not controlled by the teacher, it is not surprising to find small relationships between students' and teachers' self-reports, especially since student class averages were used as the unit of analysis, rather than individual student levels of persistence, goal orientation, and self-efficacy. These aggregate scores would not be as sensitive to the relationships as would correlations using individual students, and most likely mask influences on some students. Still, the lack more significant relationships between teachers' formative assessment practices and class averages of student levels of motivation is somewhat surprising given other research (e.g., Brookhart & Durkin, 2005; Shepard, 2000) that shows a relationship between formative assessment and achievement. The statistically significant positive relationships between overall formative practices and class averages of student motivation, suggests an association between at least some formative practices and student motivation. When teachers reported using several types of formative assessments practices, students were more likely to report higher levels of motivation.

The fact that many student-perceived teacher practices were correlated to student motivation suggests that some teacher practices are important. This includes praising hard work when wrong, stressing the importance of learning, and making written comments. The results suggest that teachers who are perceived by students as having a more serious academic orientation are more likely to have motivated students. Of course, this is a correlation, and other teaching practices could be important. For the purpose of this paper, the findings also suggest that some undesirable practices, such as pointing out

gets good grades, are not considered positive formative assessment practices, yet these are related to student motivation.

Overall, these relationships should be viewed with caution due to the lack of statistical power and likelihood of committing a Type 2 error. Also, effect sizes are small. And, with correlational data there may be other factors related to the ones studied that account for the relationships.

The study makes a contribution to our knowledge of formative assessment in several ways. First, it is important to the formative assessment literature because the study was implemented in a way that linked teachers' formative assessment practices and different types of student motivation. It is unique in having secondary students report motivation for a specific class so that those students' responses could be correlated to teachers' formative assessment practices. Second, consistent with previous research, overall formative practices showed a positive relationship with student motivation. Third, this study examined specific aspects of formative assessment, and showed that such practices as guiding further instruction and learning may be a weakness in the way teachers apply formative assessment. Fourth, this study illustrates the difficulty of researching questions related to student motivation and teachers' formative assessment practices. Last, the descriptive results have important implications for instructional practices by revealing that many or most secondary teachers do not use formative assessment practices nearly as much as what is suggested in the literature, and report that several critical components of formative assessment, especially instructional correctives, are not widely used.

Given what we know about teachers working in the current environment of high-stakes testing, it is possible that teachers perceive formative assessment practices as impractical or time-consuming. Discussions with teachers about use of formative assessment reveal that they believe they cannot afford devoting instruction time to these practices because of need to teach to standards and cover the content in tightly scheduled ways (Abrams, 2006). However, previous research supports a link between the use of formative assessment and teacher effectiveness. Furthermore, Abrams (2006) suggests that the use of formative assessment may in fact serve the goals of the high stakes testing mandate, as well as improving teaching.

Support in the literature for the use of formative assessment in schools is strong; however, in this sample teachers are not using formative assessment practices as frequently as one would expect. It could be important for future research to determine why teachers may not be using these practices. As mentioned above, teachers may perceive formative assessment as unwieldy or inefficient. However, other possible barriers to using formative assessment practices may exist. Black and Wiliams (1998) suggest that an important obstacle may be conflicts among value systems, belief systems and structures, agendas, and values in institutions. There also may be a lack of professional development for both teachers and principals. These barriers should be explored fully.

Further research needs be conducted to assess outcomes and effects of using formative assessment practices, especially to use information generated from formative assessment to change instruction for all students, to motivate students in an appropriate manner, and to provide instructional correctives to individual students. In addition to

gathering self-report data on practices, research regarding assessment practices should

evaluate whether teachers feel adequately prepared or trained in these practices (and their

purpose). Individual teachers, principals, and school systems can be evaluated on these

dimensions. Moreover, school administrators or educators who wish to increase the use

of formative assessment practices should consider these training and development issues.

References

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.

Anderman, E., & Maehr, M. L. (1994). Motivation and schooling in the middle grades. *Review of Educational Research, 64*, 287-309.

Adrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In G. J. Cizek and H. L. Andrade (Eds.), *Handbook of formative assessment* (pp. 90-105). New York, Routledge.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*, 139-148.

Brookhart, S. M. (1997). A theoretical framework for the role of classroom assessment in motivating student effort and achievement, *Applied Measurement in Education, 10*(2), 161-180.

Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record, 106*(3), 429-458.

Brookhart, S. M. (2007). Expanding views about formative assessment: A review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into Practice* (pp. 43-62). New York: Teachers College Press.

Brookhart, S. M . & Durkin, D. T. (2003). Classroom assessment, student motivation, and achievement in high school social studies classes. *Applied Measurement in Education, 16*, 27-54.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In G. J. Cizek and H. L. Andrade (Eds.), *Handbook of formative assessment* (pp. 3-17). New York, Routledge.

Finn, J. D., & Cox, D. (1992). Participation and withdrawal among fourth-grade pupils. *American Educational Research Journal*, *(29)*1, 141-162.

Guskey, T. R. (2010). Formative assessment: The contributions of Benjamin S. Bloom. In G. J. Cizek and H. L. Andrade (Eds.), *Handbook of formative assessment* (pp. 106-124). New York, Routledge.

McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

McMillan, J. H. (2002). Elementary school teachers' classroom assessment and grading practices. *Journal of Educational Research, 95*(4), 203-214.

McMillan, J.H. (2007). Formative classroom assessment: The key to improving student achievement. In J. H. McMillan (Ed.), *Formative classroom assessment: Theory into practice* (pp. 1-7). New York: Teachers College Press.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., & Urdan, T., (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*, Ann Arbor, MI: University of Michigan.

Pintrich, P.R, & Schunk, D. H. (2002). *Motivation in education: Theory, research, and applications* (3rd Ed.). Englewood Cliffs, NJ: Prentice Hall.

Shepard, L. A. (2000) The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership, 63*(3), 66-70.

Wiliam, D. (2010). An integrative summary of the research literature and implicas for a new theory of formative assessment. In G. J. Cizek and H. L. Andrade (Eds.), *Handbook of formative assessment* (pp. 18-40). New York, Routledge.

Wiliam, D., & Leahy, S. (2007). A theoretical foundation for formative assessment. In J. H. McMillan (Ed.), *Formative Classroom assessment: Theory into practice* (pp. 29-42). New York: Teachers College Press.

Zimmerman, B. (2000). Self-Efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82-91.

Table 1
Items and Factor Loadings

Item	Factor loadings			
	1	2	3	4
It's important to me that other students in my class think I am good at my class work.	**.64**	.20	.08	.10
I don't want my teacher to think I know less than other students in class.	**.49**	.12	.04	.13
One of my goals is to show others that class work is easy for me.	**.72**	.14	.08	-.05
It's important to me that I don't look confused in this class.	**.73**	.03	.06	.03
It's important to me that I look smart compared to other students in my class.	**.82**	0	.09	-.01
I don't want anyone in the class to know if I'm having trouble doing the work.	**.66**	-.04	-.07	0
One of my goals is to look smart compared to other students in this class.	**.82**	-.02	.06	-.03
One of my goals is to show others that I'm good at my class work.	**.76**	.17	.11	-.01
It's important to me that I learn a lot in this class.	.08	**.75**	.16	.15
It's important to me that I completely understand my class work.	.12	**.57**	.27	.24
One of my goals is to learn as much as I can in this class.	.08	**.81**	.22	.15
It's important to me that I improve my skills in this class.	.12	**.78**	.06	.07
I am certain that I can understand the ideas taught in this class.	.03	.04	**.67**	.26
I expect to do well in this class.	.10	.38	**.66**	.11
I'm sure I can do an excellent job on the work in this class.	.11	.19	**.76**	.14
I know that I will be able to learn the material for this class.	.06	.19	**.75**	.17
I think that I will receive a good grade in this class.	.07	.14	**.78**	.09
If I can't understand an assignment at first, I keep going over it until I understand it.	.04	.21	.08	**.66**
I would rather do assignments that are challenging than assignments that are really easy.	.04	.20	.18	**.53**
If I don't know the answer to a question, I try to figure it out on my own.	.08	.03	.12	**.58**
I stop trying when the assignment or homework is very difficult. (R)	.00	.11	.11	**.69**
I give up when I make mistakes. (R)	-.05	.06	.13	**.62**

Table 2
Subscale Reliabilities

Component	Subscale	Cronbach's Alpha
1	Performance	.86
2	Mastery	.80
3	Efficacy	.82
4	Persistence	.65

Table 3.
Frequencies, Percentages, Means, and Standard Deviations of Teacher Formative Assessment Practices.

Item	Not at all		Minimally		Some		Quite a bit		Extensively		Mean	SD
	f	(%)	*f*	(%)	*f*	(%)	*f*	(%)	*f*	(%)		
Feedback (written or verbal) on performance that was given privately to each student	9	(6)	24	(16)	53	(34)	45	(29)	24	(16)	3.33	1.09
Specific, individualized feedback (written or verbal)	5	(3)	20	(13)	52	(33)	52	(33)	27	(17)	3.49	1.03
Feedback (written or oral) that contained suggestions for further learning	8	(5)	21	(13)	63	(40)	52	(33)	12	(8)	3.25	0.96
Formative assessments (i.e., assessment given during instruction to check student learning; anecdotal or structured)	3	(2)	15	(10)	44	(29)	65	(43)	26	(17)	3.63	0.95
Assessments that diagnosed student weaknesses (e.g., pretest; diagnostic assignment or questions)	13	(8)	52	(33)	55	(35)	30	(19)	6	(4)	2.77	0.98
Assessments that were used to guide further instruction	10	(7)	24	(16)	71	(46)	38	(25)	10	(7)	3.09	0.96
Ongoing, in class monitoring of student progress	0	(0)	10	(7)	45	(29)	66	(43)	29	(19)	3.79	0.85

Table 4.

Correlations Between Student Subscales and Between Subscales and Formative Assessment Items

	Performance	Mastery	Efficacy	Persistence
Performance Subscale	1			
Mastery Subscale	.44**	1		
Efficacy Subscale	.36**	.48*	1	
Persistence Subscale	.18*	.38**	.45**	1
Feedback given privately	-.03	.08	.06	-.04
Specific, individualized feedback	.01	-.04	-.03	-.05
Feedback containing suggestions for further learning	-.00	.08	.09	-.01
Formative assessment	-.06	.11	.17*	.05
Assessment to diagnose student weaknesses	.11	.19*	.17*	.08
Assessment used to guide further instruction	.01	.13	.07	.08
Ongoing, in-class monitoring of progress	.04	.10	.16	.08
Total of formative items	.17*	.30**	.22**	.20**

** Correlation is significant at the .01 level; * Correlation is significant at the .05 level

Table 5.

Correlations Between Student Motivation Subscales and Student-Reported Teacher Practices

	Performance	Mastery	Efficacy	Persistence
Teacher thinks mistakes are ok as long as we learn	-.06	.13	.10	.03
Teacher points out those who get good grades	.35**	.29**	.22*	.02
Teacher tells us it is important to learn for SOLs	.40**	.55**	.23**	.13
Teacher has us keep track of own grades	.29**	.17*	.11	.17*
Teacher praises trying hard even when wrong	.18*	.39**	.29**	.08
Teacher tells us who gets highest scores	.22**	.16	.14	-.01
Teacher tells us it is important to learn so we understand	.17*	.52**	.26**	.29**
Teacher's written comments help me understand	.17*	.48**	.34**	.25**
Teacher tells us how our grades compare to peers	.20*	-.09	.03	-.15

** Correlation is significant at the .01 level; * Correlation is significant at the .05 level