

THE FORMATIVE USE OF SUMMATIVE ASSESSMENTS AT THE U.S.
ARMY COMMAND AND GENERAL STAFF SCHOOL: A QUALITATIVE
CASE STUDY

by

DENNIS S. BURKET

B.S., Troy University, 1986
M.S., Webster University, 1995
M.B.A., Webster University, 1996

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF EDUCATION

Department of Educational Leadership
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Abstract

This qualitative case study sought to understand in what formative ways instructors in one teaching department of the U.S. Army Command and General Staff College (CGSC) used common summative assessments and what similar practices instructors used as a result of common summative assessments. This research analyzed data from semi-structured interviews with purposefully selected participants, instructors in the Department of Army Tactics at CGSC, a representative mixture of civilian and active duty. This research confirmed that the formative use of summative assessments was typical among Department of Army tactics instructors and that continued or expanded formative use of summative assessments will increase student learning. Because so much of assessment is context dependent, this research will add to the body of knowledge in a particular area that the current literature did not fully address; the formative use of common summative assessments in higher education.

Four conclusions were drawn from analysis of this research. First, the formative use of common summative assessments, especially feedback given to students, was typical of the Department of Army Tactics instructor, essentially a normal part of the assessment process. Second, DTAC instructors did not have a common understanding of the difference between summative and formative assessment, how they used the information gathered was more important than what the instrument was labeled. Third, “teaching to the curriculum” instead of “teaching the test” was typical in DTAC, an indication that the instructors saw their role beyond just preparing students for upcoming assessments. Fourth, the stratification of students during the grading process was typical, with the unintended consequence of students not being judged on quality of work alone.

THE FORMATIVE USE OF SUMMATIVE ASSESSMENTS AT THE U.S.
ARMY COMMAND AND GENERAL STAFF SCHOOL: A QUALITATIVE
CASE STUDY

by

DENNIS S. BURKET

B.S., Troy University, 1986
M.S., Webster University, 1995
M.B.A., Webster University, 1996

A DISSERTATION

Submitted in partial fulfillment of the requirements for the degree

DOCTOR OF EDUCATION

Department of Educational Leadership
College of Education

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2016

Approved by:

Major Professor
Dr. Sarah Jane Fishback

Copyright

DENNIS SAMUEL BURKET

2016

Abstract

This qualitative case study sought to understand in what formative ways instructors in one teaching department of the U.S. Army Command and General Staff College (CGSC) used common summative assessments and what similar practices instructors were used as a result of common summative assessments. This research analyzed data from semi-structured interviews with purposefully selected participants, instructors in the Department of Army Tactics at CGSC, a representative mixture of civilian and active duty. This research confirmed that the formative use of summative assessments was typical among Department of Army tactics instructors and that continued or expanded formative use of summative assessments will increase student learning. Because so much of assessment is context dependent, this research will add to the body of knowledge in a particular area that the current literature did not fully address; the formative use of common summative assessments in higher education.

Four conclusions were drawn from analysis of this research. First, the formative use of common summative assessments, especially feedback given to students, was typical of the Department of Army Tactics instructor, essentially a normal part of the assessment process. Second, DTAC instructors did not have a common understanding of the difference between summative and formative assessment, how they used the information gathered was more important than what the instrument was labeled. Third, “teaching to the curriculum” instead of “teaching the test” was typical in DTAC, an indication that the instructors saw their role beyond just preparing students for upcoming assessments. Fourth, the stratification of students during the grading process was typical, with the unintended consequence of students not being judged on quality of work alone.

Table of Contents

List of Figures	x
List of Tables	xi
Acknowledgements.....	xii
Dedication.....	xiii
Chapter 1 - Introduction.....	1
Overview of the Issues.....	1
Background.....	3
Formative Use of Summative Assessment	3
Formative Use of Feedback	4
Impact of Assessments on Motivation.....	5
Statement of the Problem.....	6
Purpose of the Study.....	7
Research Questions.....	8
Brief Description of Methodology.....	8
Case Study Context.....	9
Significance of the Study.....	9
Limitations of the Study	10
Assumptions.....	10
Human Rights and Ethical Standards	11
Definitions of Terms.....	11
Summary.....	13
Chapter 2 - Literature Review.....	14
Background of Assessment.....	14
History of Assessment	14
Learning Theories and Assessment.....	16
The Assessment Process	18
Contemporary Assessment.....	20
Summative Assessment	22
Formative Assessment	23

Assessment and Learning	26
Formative Use of Summative Assessment	29
Balanced Assessment Model	35
Common Assessments	35
Embedded and Add-on Assessments	36
Progressive and Continuous Assessment.....	36
EPPI-Centre Studies (2002 and 2004)	37
CGSS Assessment Process	39
Using Feedback to Adjust Teaching.....	41
Using Grades to Change Student Motivation to Learn.....	44
Norm-Referenced and Criterion-Referenced Approaches	46
Standardized Tests	47
Teaching to the Test.....	49
The Grading Process	49
Goal Orientations	51
Summary.....	51
Chapter 3 - Methodology.....	53
Research Questions.....	53
Research Design	53
Setting.....	54
Population	57
Sampling Strategy.....	58
Purposeful Sample	59
Procedures.....	62
Data Collection	62
Focus Group.....	63
Pilot Interviews	64
Semi-structured Interviews	64
Data Analysis.....	65
Trustworthiness.....	66
Role and Background of the Researcher.....	67

Ethical Considerations	69
Summary	69
Chapter 4 - Findings.....	71
Introduction.....	71
Qualitative Methodology	71
Demographics	71
Participant Profiles.....	72
Qualitative Interview Results.....	75
Theme One: Feedback Given to Students.....	76
Written feedback was provided to students	77
Common errors/trends of the group discussed or provided to students.....	80
Feedback applicable to other assignments	82
Use of exemplars.....	85
Individual coaching.....	87
Timeliness of feedback	89
References provided for further learning.....	91
Theme Two: Modification of Teaching and Learning Activities	92
Recommended adjustments to the common summative assessment instrument after the assessment was given.....	93
Adjustments to learning activities or teaching style after the assessment is given.....	97
Adjusting learning activities or teaching style before the assessment is given	102
Reflection by instructors about the need to modify teaching or learning activities after the assessment is given	105
Theme Three: Preparation of Students for Upcoming Summative Assessments	108
“Teaching to the test” vs. “Teaching the test” (Learning Objectives vs. test questions)	108
Practice sessions for students.....	116
Use of examples – ‘what right looks like’	117
Theme Four: Use of the Grading Process	118
Student motivation – grades/prioritization.....	119
Stratification of students – diverse population.....	127
Calibration amongst teaching team members	136

Summary.....	139
Chapter 5 - Analysis, Discussion, Implications, and Recommendations	140
Introduction.....	140
Restatement of the Problem	140
Restatement of the Purpose.....	140
Research Questions in Review.....	141
Discussion of the Study	141
Research Question One.....	142
Research Question Two	145
Implications for Military Education	149
Implications for Adult Education	149
Recommendations.....	149
Recommendations for Future Research.....	150
Reflections	150
Concluding Remarks.....	151
References.....	152
Appendix A - Interview Protocol.....	166
Appendix B - Informed Consent.....	168
Appendix C - Instructor Survey Form	170
Appendix D - Common Assessment Classification Worksheet	171
Appendix E - Non-disclosure Form.....	172
Appendix F - Kansas State University IRB Approval.....	173
Appendix G - CGSC IRB Approval	174

List of Figures

Figure 2.1 Teaching, Learning, and Assessment as a Continuous Four-Step Cycle	19
Figure 2.2 Assessment for Summative Purposes	23
Figure 2.3 Assessment for formative purposes	25
Figure 2.4 A possible dimension of assessment purposes and practices	31
Figure 2.5 Accelerated Adult Education Version of Types of Formative Assessment	33
Figure 3.1 CGSC Classroom at the Lewis and Clark Center.....	56
Figure 3.2 Sequential Flow of this Qualitative Case Study	63

List of Tables

Table 2.1 Compilation of Assessment Characteristics	28
Table 3.1 Population Characteristics: All DTAC Instructors (AY 2013).....	58
Table 3.2 Participant Characteristics: DTAC Instructors at Ft Leavenworth (AY2012-2013)	61

Acknowledgements

I would like to acknowledge the U.S. Army Command and General Staff College, specifically, the leadership and members of the Department of Army Tactics (DTAC) for fully supporting my research and providing needed resources. I would like to thank the study participants, all members of DTAC, for their willingness to participate and provide honest and candid responses to the interview questions about their experiences. I would like to thank Dr. Kevin Shea, Dr. Bill McCullum, and Dr. Shawn Cupp for their contributions and assistance and Dr. Jack Kem who provided a critical eye and detailed feedback that help refine my interpretation of the research.

I would like to acknowledge the Kansas State University faculty and committee members who contributed their time and expertise to my endeavor and never wavered in their support. Dr. Jane Fishback, my major professor, who provided mentorship, guidance, and direction and never once doubted that I would finish. Dr. Royce-Ann Collins, for her insights, expertise, and willingness to support at any time. Dr. Jim Martin and Dr. Jana Fallin for their assistance, guidance and timely feedback.

Finally, I would like to acknowledge and thank my wife, Susanne, who provided the encouragement and support I needed to finish this project and never once doubted that I would finish.

Dedication

This dissertation is dedicated to my wife, Susanne, and daughters, Katrina and Alyssa who supported me throughout my research and were understanding of the time I spent time away from family events - I'm finally done!

Chapter 1 - Introduction

Overview of the Issues

The increasing focus on assessment of student learning influences most institutions of higher education in the United States today (Banta, et al., 2009). Institutions risk losing accreditation, resources, or applicants if they fail to comply with stated requirements (Ewell, 2008; Suskie, 2009). Teachers fearing the loss of employment or promotion opportunities change the focus of their teaching in order to ensure acceptable results on student assessments. For students, individual assessment shapes experiences and influences behavior more than the teaching they receive and can mean the difference between success or failure (Bloxham & Boyd, 2007). Because of the effect assessment can have on students, “there is more leverage to improve teaching through changing assessment than there is in changing anything else” (Gibbs & Simpson, 2004, p. 22).

Assessment similarly affects United States military institutions of higher learning which are charged with developing the future leaders of our nation’s military (*CJCS, 2012*). In addition to meeting accreditation requirements, these military institutions, such as the U.S. Army Command and General Staff College (CGSC) located at Fort Leavenworth, Kansas, must formally assess each student officer to ensure that they have met required learning objectives. Unlike most civilian institutions, the students who attend the Command and General Staff College can expect to be deployed worldwide and lead soldiers in dangerous and uncertain environments (U.S. House of Representatives Report, 2010; Army Learning Concept for 2015, 2010), sometimes within months of graduation.

In times of uncertainty in the world, such as now, our military has traditionally relied on education to prepare officers to meet unknown future challenges (*CJCS, 2012*). The Command and General Staff College leadership is charged with implementing an adaptive curriculum that can meet changes to the environments where its graduates will operate. CGSC continually reevaluates its curriculum with a focus on identifying the best mixture of content and delivery methods for classes beginning each academic year (Davis & Martin, 2012).

To meet these demands, the Command and General Staff College developed a curriculum that employed a large number of summative individual and group assessments of different types

designed to yield data that will be used to make decisions about individuals and courses (*CGSC Self Study Report, 2005*). Ultimately, the responsibility for evaluating these student officers falls on the individual instructor in the classroom who is charged with the dual, sometimes competing tasks of evaluating students, often referred to as summative assessment and developing student learning, often referred to as formative assessment (Ewell, 2002).

Although assessment is of central importance in education (Taras, 2005), there is a lack of commonality in the related definitions and terminology. In primary, secondary, and higher education, assessment has many meanings that are contextual in nature. Assessment is an ongoing process that can occur before, during, and/or after instruction is delivered. Assessment can focus on individual student learning, on an entire class of students, a program, the educational institution, or an educational system as a whole.

The terms summative and formative assessment are fundamental to understanding assessment in education. Summative assessment focuses on summing up or summarizing student achievement (Bloom, Hastings, & Madus, 1971; Sadler, 1989; Shavelson, 2006), while formative assessment is centered on active feedback that assists learning (Black & Wiliam, 2004; Sadler, 1989; Shavelson, 2006). Many educators refer to summative assessment as assessment *of* learning and formative assessment as assessment *for* learning (Black & Wiliam, 2003; Broadfoot, 2008; Stiggins, 2002). The use of summative assessment for formative purposes, the focus of this study, was an area of assessment often described as underused, but with much potential for improving student learning (Black, Harrison, Lee, Marshall & Wiliam, 2004; Stiggins, 2009; Taras, 2008).

Although there is a growing base of assessment literature that instructors can consult, most of the supporting research is from the primary and secondary education levels in the U.S. and Britain (Martin & Collins, 2011). Educational practice, at all institutions, is contextual in nature and activities need to be adapted to the classroom environment (Laurillard, 2002; Suskie, 2009; Walvoord & Anderson, 2010) meaning the responsibility falls on instructors and teaching teams to “critically evaluate assessment processes and procedures as they are used and developed within their local context” (Bloxham & Boyd, 2007, p. 7). Linda Suskie (2009) further explained the contextual nature of assessment, “While it is systematic, it is context specific rather than generalizable, informal rather than rigorous, and designed to inform individual rather than general practice” (p. 60).

The assessment strategy of a particular course has a major impact on student activity (Snyder, 1973). Measuring a desired outcome will do little to improve it without examining the processes that led to the outcome (Banta, Black & Jones, 2009). Bloxham and Boyd (2007) argued that instructors, like those at the US Army CGSC, who teach approved curriculums with fixed assessments “can have a significant influence on the assessment strategy of the course and how students perceive and engage in assessed tasks” (p. 49). For the purpose of informing practice, this qualitative case study sought to understand, from the instructor perspective, the effect summative assessment had on the activities instructors used to improve student learning.

Background

Formative assessment has an extensive research base that draws on both cognitive and motivational research (Shepard, 2005), but few studies have been conducted with students in higher education (Yorke, 2008). This study was supported by the literature associated with summative assessment used for formative purposes, formative use of feedback, and student motivation and is positioned within the activities individual instructors use while teaching a standardized (uniform) curriculum with common assessments of learning.

Writing about curriculum, Michael Scriven (1967) recognized the need to distinguish between what he saw as the dual roles of evaluation. He observed that while evaluation “may have a role in the on-going improvement of the curriculum” (Scriven, 1967, p. 41), it could also be used to examine the curriculum of an entire program (William, 2006). Based on this recognition, he introduced the terms ‘formative evaluation’ and ‘summative evaluation’ to differentiate these distinct roles (Scriven, 1967, p. 43). Two years later, Benjamin Bloom (1969) suggested that the distinctions Scriven applied to curriculum could also be applied to the evaluation of student learning, or what is commonly referred to today in the United States as ‘assessment’ (Allal & Lopez, 2005; Knight, 1995; William, 2006).

Formative Use of Summative Assessment

Although a large body of the assessment literature today aims to delineate the differences between summative and formative assessment: Bell and Cowie (2001); Black et al. (2003); Bloom (1969); Broadfoot and Black (2004); Dunn and Mulvenon (2009); and Scriven (2006) agreed that the same tests could be used for both summative and formative purposes. An

increasing volume of literature advocates use of the same assessments for both summative and formative purposes (Burke, 2010; Stiggins & Arter, 2009; Taras, 2008).

Many educators asserted that summative assessment happens too far along in the learning process to make instructional adjustments (Ainsworth & Viegut, 2006; Burke, 2010; Garrison & Ehringhaus, 2011) but research does not reflect this. Black et al. (2003a) adjusted their ideas on using formative and summative assessment to support each other in the classroom context because teachers refused to separate them (Black et al., 2004, p. 31; Taras, 2005) and research evidence showed that teachers found it useful to use summative assessment for formative purposes (Black et al., 2004, pp. 55-56, Taras, 2001, p. 610).

Based on a study of 50 education lecturers (instructors) at an English university, Taras (2008) posited that due to problems of inconsistencies in the literature, lecturers did not understand the relationship between summative and formative assessments. Responses to survey questions indicated that lecturers were not sure about the relationship of formative assessment to grading and the role of summative assessment when it is not the 'final' summation. She concluded her study with several recommendations for future study including: "Future research needs to ask how we can reconcile formative and summative assessment so that they are mutually supportive: examining the processes of assessment seems one possible way of doing this" (Taras, 2008, p. 189).

Formative Use of Feedback

Most often associated with formative assessment, feedback has been shown to help low achievers more than others (Black et al., 2003a; Harlen, 2004b). Taras (2010) explained the relationship of feedback to assessments: "whereas summative assessment produces feedback, formative assessment must use feedback" (p.10). In higher education, most exams are created internally, by individual instructors or departments within the school. Taras (2009) asserted that most feedback in higher education comes from graded work and to not link formative assessment with summative assessment in this context risks losing "the most powerful and central learning tool [higher education] has" (p. 66).

Wlodkowski (2008) offered that "feedback is probably the most powerful communication that instructors can regularly use to affect learners' competence" (p. 315). Hattie and Timperley (2007) concurred and added that most current assessments were not effective because they

provided minimal feedback, usually in the form of grades that didn't help students understand what they needed to do to improve. Although most of the research on feedback has been at the primary and secondary levels (Hattie and Timperley, 2007), it has been shown to be critical for enhancing practice and deepening adult learning (Keeton, Sheckley, & Griggs, 2002).

With the publishing of "Inside the Black Box" in *Phi Delta Kappan*, Paul Black and Dylan Wiliam spread the interest in formative assessment from its initial base in the United Kingdom to the United States (Brookhart, 2004, p. 449). This influential review of over 250 articles is widely cited as proof that using formative assessment in the classroom improves student learning (Black & Wiliam 1998a; Harlen, 2005). In this same article, Black and Wiliam proposed a definition of assessment that follows Ramaprasad (1983) and Sadler's (1989) central use of feedback and connects it to the adaptation of activities by teachers to improve student learning:

We use the general term assessment to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student standards (Black & Wiliam, 1998b, p. 140).

Impact of Assessments on Motivation

The literature on motivation acknowledged the relationship between graded assessments and student motivation (Laurillard, 1997; Ramsden, 1992; Rust, 2002; Snyder, 1973). Walvoord and Anderson (2010) noted that grading affects how students study, what they focus on, how much time they spend, and how involved they become in a course. Whereas grades are often used to control student behavior (Rowntree, 1987), Black et al (2004), felt that a formative approach can counteract student's obsession with grades and redirect their interest towards learning. If students begin to see summative assessment as formative, and see it as being good, they will accept it more. Learning-oriented students are more engaged and develop a deeper understanding of the subject matter they are studying (Bloxham & Boyd, 2007).

In 2002 and 2004, The EPPI-Centre (Evidence for Policy and Practice: *Information and Co-ordinating Centre*), part of the Social Science Research Institute of Education, University of London, conducted a review of research on the effects of the use of summative and formative

assessment for the same purpose (Harlen, 2005). Looking at the impact summative assessment had on student motivation for learning, Harlen and Crick found that when preparing students to pass high-stakes tests was the focus, teachers resorted to a transmission style of teaching (Black & Wiliam, 1998a; Linn 2000; Stiggins, 1999), and that students could be trained to pass any type of test, even ones designed to assess higher levels of thinking (Harlen, 2005). When teachers participated in the development of criteria used on summative assessments, they were more able to use them reliably when evaluating students' work (Hargreaves et al., 1996; Harlen, 2004a). Looking at assessments used for internal purposes, there was evidence that feedback from earlier assessments impacted the amount of effort students applied for future tasks of the same kind (Brookhart & DeVoge, 1999; Carter, 1997/8, Harlen, 2004b). There was also evidence that teachers changing assessment processes, practices and explanations they gave to students could lead to better student learning (Flexer et al., 1995; Harlen, 2004b).

Walvoord and Anderson (2010) made no distinction between formative and summative assessment arguing instead that “grading infuses everything that happens in the classroom” (p.1) and that instead of trying to pretend that grading doesn't affect students, educators should “use its power for student learning” (p. 1). They described grading as a complex context-dependent process by which a teacher uses classroom tests and assignments to shape student motivation with the aid of feedback and adjustments to teaching methods (Walvoord & Anderson, 1998; 2010).

The literature on assessment theory supports the use of summative assessment for formative purposes. Research at the primary and secondary levels showed that the use of graded assessments effected activities instructors used with both positive and negative effects on student learning. Literature on how the use of summative assessments in higher education effects activities instructors chose to use to develop student learning is not fully developed.

Statement of the Problem

Because assessment of student learning is context-dependent (Walvoord & Anderson, 2010), previous literature does not fully address, from the higher education instructor's perspective, the effect that using common graded assessment instruments has on the activities used to develop student learning. Instructors who teach standardized curriculums that use common assessments of learning need methods to develop student learning that are effective in

their environment, where the lack of control over scheduling and curriculum limits the practices they can employ (Bloxham & Boyd, 2007).

The effect of graded assessment must be understood to address the external requirement for rigor in the curriculum. In 1989, The House Armed Services Committee Panel on Primary Military Education (PME) led by Representative Ike Skelton established the requirement for graded activities (*U.S. House of Representatives Committee on Armed Services*, 2010) arguing that the “sine qua non of a PME school’s rigor is graded activities” (p. 272).

The *Command and General Staff College Self-Study Report* (2005) submitted to the Higher Learning Commission of the North Central Association of Colleges and Schools for accreditation purposes identified assessment methodology as a concern. The report suggested the need for a more detailed understanding of current faculty assessment processes before CGSS undergoes accreditation again. There are more current CGSC reports, but they were not available during the time period participants were interviewed and so were not used.

Findings from the 2010 Command and General Staff School Faculty Survey provided further evidence of the need to understand how Department of Army Tactics (DTAC) instructors use common summative assessment instruments to assess student learning. This survey found instructors felt that students were not accurately assessed and that curriculum assessments did not accurately measure student learning (*CGSS Faculty Survey for AY 2010*). It was beyond the scope of the 2010 Faculty Survey to determine if the activities instructors chose to use affect their perception of the accuracy of assessment instruments.

Developers of common assessment instruments needed to fully understand, from the instructor’s perspective, the effects these assessments can have on student learning. In an environment with a large number instructors and constant turnover, there is a need for sharing information about practices that might help others improve student learning in their classrooms.

Purpose of the Study

The purpose of this exploratory case study was to understand, from the Department of Army Tactics instructor’s perspective; were there formative ways common summative assessments were used to improve student learning and what similar practices, if any, instructors employed as a result of using common summative assessments of student learning.

Research Questions

This study sought to answer two research questions:

Research Question One.

Are there formative ways Department of Army Tactics instructors use common summative assessments?

Research Question Two.

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

The research questions were used to guide semi-structured interviews with participants. The full list of interview questions is located in Appendix E, Interview Protocol.

Brief Description of Methodology

This study used a qualitative bounded case study methodology. Creswell explained that case study research involves the in-depth exploration of an activity or process, and individuals in a bounded system (2009). This study was bounded by a one-month period of data collection (June 2013), with participants from the Department of Army Tactics, part of the Command and General Staff College, located at Fort Leavenworth, Kansas. Purposive sampling was used to identify appropriate “information-rich” (Patton, 1990) military and civilian study participants. Selected participants had taught the DTAC Intermediate Level Education Common Core Curriculum at least once, had been a lesson or course author, and a Staff Group Advisor. Rationale for each of these criterion are described in Table 3.2.

Data collection activities consisted of a focus group interview, two pilot interviews and ten research interviews. The primary means of data collection was semi-structured interviews with open-ended questions to capture the experiences of the study participants in their own words. Information saturation was reached with the chosen participants, allowing research to continue without the need to select more participants. Interviews were tape recorded and professionally transcribed. Data analysis was based on procedures designed to produce credibility, dependability, confirmability, and transferability (Guba & Lincoln, 1985). Interview

transcripts were member checked for accuracy and precision. Episodic peer debriefing involving several expert peers was used to enhance the accuracy of the analysis (Creswell, 2009).

Participants signed statements advising them of their rights. Individuals who transcribed or reviewed data signed non-disclosure statements. All data, records, and field notes were and continue to be safeguarded to prevent public disclosure of interview responses.

Case Study Context

To help the reader understand the context of this study, this section briefly describes the setting in which the research was conducted. The study location was the U.S. Army Command and General Staff College (CGSC), located at Fort Leavenworth, Kansas. The CGSC is accredited by the Higher Learning Commission (HLC) of the North Central Association of Colleges and Schools (NCACS) as a Master's Degree Granting Institution. The College also sustains accreditation credentials to deliver Joint Professional Military Education (JPME) and Army Professional Military Education (PME). As an institution accredited by the North Central Association of Colleges and Schools, CGSC subscribes to the American Association of University Professors 1940 Statement on Academic Freedom (*CGSC Catalog*, 2012). As of March 2015, CGSC is part of the Army University. CGSC encourages free thought and the free flow of ideas and encourages aggressive examination of all academic subjects.

The instructor participants all taught the 10-month Command and General Staff Officer's Course (CGSOC), a graduate-level course that provides intermediate professional education for selected mid-career officers (majors and lieutenant colonels) from all military services of the United States as well as over 90 countries around the world (*CGSC Self-Study Report*, 2005). In recent years, students from 15 other US Government Agencies such as the Department of State, National Geospatial and Intelligence Agency, Federal Bureau of Investigation, Border Patrol, and the Department of Homeland Security with limited military background have also been attending.

Significance of the Study

Existing research on the effect of summative assessments on instructor activities was focused at the primary and secondary levels because of the need to prepare students for high-stakes testing at district and state levels. Higher education does not have this requirement, but there is still a need to understand how summative assessment can be used to develop student

learning. An increasing amount of literature advocates the use of summative assessment for formative purposes, but it is unclear how this affects the activities of instructors in higher education and more specifically in a seminar environment.

The contextual nature of education necessitates that studies be local to be most effective (Suskie, 2009; Walvoord & Anderson, 2010). At the time of data collection (June 2013), the Department of Army Tactics had an instructor population of 122 with a 10 % annual turn-over rate. By gaining thorough descriptions of how the use of common assessments of student learning affect the activities instructors use to develop student learning, this exploratory case study provides the Command and General Staff School faculty and staff an opportunity to improve curriculum development, teaching practices and student learning. This research contributes to the body of knowledge of assessment and the fields of adult and higher education. More broadly, this research also contributes to Primary Military Education (PME), and Joint Primary Military Education (JPME).

Limitations of the Study

The following limitations apply to this research:

1. The results of this research were dependent upon the ability of the interviewer to remain focused, unbiased, and objective.
2. The selection of study participants overrode concern for equal representation with respect to age, gender or ethnicity.
3. DTAC instructors at satellite locations were not interviewed because of the cost of travel and the desire to conduct face-to-face interviews versus Video Tele Conferencing.
4. Participants were resident DTAC faculty from the Fort Leavenworth, Kansas campus only. Because this is a unique population, the study findings may be limited to resident CGSOC instruction.
5. The results of the study were limited by the accuracy and truthfulness of the participants' responses to interview questions.

Assumptions

The following assumptions were made for the purpose of this research:

1. Faculty members are intentional, to varying degrees, about the ways by which they attempt to develop student learning.

2. The study participants will provide honest and candid responses to the interview questions and truthfully represent their experiences with common assessments of student learning in the CGSOC.
3. The peer relationship of the primary researcher to individual study participants will establish credibility and trust with individual study participants.

Human Rights and Ethical Standards

This research was conducted in compliance with Kansas State University (KSU) policy for research with human subjects. The committee on Research Involving Human Subjects and Institutional Review Board (IRB) administers this program in accordance with *Kansas State University Committee for Research Involving Human Subjects Institutional Review Board (IRB) 2012*.

Due to participants being faculty at the U.S. Army Command and General Staff College (CGSC), this research was also compliant with U.S. Army Command and General Staff College requirements for research with human subjects. The Quality Assurance Office (QAO) of CGSC implements U.S. Department of Defense regulations and instructions, and Army regulations as directed in CGSC Bulletin No. 40, *Research within the GCSC* (2010), which describes the CGSC Institutional Review Board (IRB) process.

Definitions of Terms

The following definitions were used for the purposes of this study:

Assessment. A process that includes all the activities undertaken by teachers that provide information to be used as feedback to modify teaching and learning activities (adapted from Black & Wiliam, 1998b).

Assessment of Student Learning. A process of documenting student achievement of learning objectives. Assessment is defined in measurable terms gathered using rubrics, classroom assessment techniques, examinations, etc. (U.S. Army CAC, LD&E Bulletin No. 30, 2008).

Command and General Staff College (CGSC). A graduate-level education institution that provides intermediate professional education for mid-career and senior officers from all military services as well as approximately 90 countries around the world. The college is comprised of five separate schools, each of which focused on a specific component of professional military education (Adapted from the U.S. Army CGSC Self Study Report, 2005).

Command and General Staff School (CGSS). The largest of the CGSC schools responsible for the education of approximately 1300 officer students each year at the Fort Leavenworth, KS campus.

Command and General Staff Officer Course (CGSOC). The ten-month curriculum administered at Ft. Leavenworth, KS in which student officers participate.

Course or Lesson Author. An instructor who is a graduate of the CGSC Faculty Development Program 3 and has been assigned the responsibility to coordinate and produce curriculum and graded assessment instruments for a specific block of instruction.

Common Assessments. “A formative or summative assessment developed collaboratively by faculty members and approved at the teaching department level that is used by an instructor in the same teaching department to evaluate (grade) student performance” (Stiggins & DuFour, 2009, p. 641).

Department of Army Tactics (DTAC). One of six teaching departments within the CGSS. The DTAC provided a program of instruction in the planning and execution of Army operations in the joint environment.

Evaluation. A systematic formalized process of gathering and analyzing data (usually both qualitative and quantitative) to determine the merit, worth, and significance of the program. Program evaluation determines whether the course supports the learning outcomes.

Faculty. “Personnel (military and civilian) who prepared, or designed professional military education (PME) curriculum, or conducted research related to PME” (Chairman Joint Chiefs of Staff, 2009, p. B-4).

Feedback. Communication from instructors to students about how they did on a graded assessment - in the form of grades or comments, verbal or written.

Formative Assessment. Assessments that provided feedback to the instructor to be used to modify teaching and learning activities, or provide feedback to the student to be used to improve learning behaviors.

Intermediate Level Education (ILE). The ten-month CGSS curriculum that consisted of the Common Core and Advanced Operations Course (AOC).

Instructor. The faculty member in a CGSS classroom who was responsible for teaching the approved curriculum. Included active duty military and Department of the Army Civilians in the academic ranks of instructor, assistant professor, associate professor, and professor.

Intermediate Level Education (ILE) Common Core Course. The course of instruction that spanned the first three months of the CGSC curriculum and provided the foundation for more advanced studies in joint, operational, and regional topics that followed in the Ft. Leavenworth 10-month course (CGSC Self Study Report, 2005). Administered at Ft Leavenworth and at each of four satellite campuses (Ft. Lee, VA; Ft. Belvoir, VA; Ft. Gordon, GA; and Redstone Arsenal, AL).

Staff Group Advisor (SGA). A faculty member assigned responsibility for the day-to-day execution of the teaching schedule within a student staff group. The SGA coached students and advised them regarding academic progress.

Summative Assessment. Assessments that measure what students have learned at a chosen point in the curriculum against a standard.

Summary

This study sought to understand if there were formative ways instructors used common summative assessments while teaching a common curriculum. This research assisted in obtaining a picture of why instructors chose to use specific activities and what similar practices, instructors identified having used as a result of using common summative assessments. The literature on assessment theory supported use of the same assessment for both summative and formative purposes. Research at mostly the primary and secondary levels showed that the use of summative assessments affected activities instructors chose to use in an attempt to improve student learning and literature on the use of graded assessments in higher education was not fully developed.

Chapter 2 - Literature Review

The purpose of this exploratory case study was to understand, from the Department of Army Tactics instructor's perspective; if there were formative ways common summative assessments were used to improve student learning and what similar practices, if any, instructors employ as a result of using common summative assessments of student learning. This study was grounded in the theory of Formative Assessment as first put forth by Paul Black, Benjamin Bloom, Arkalgud Ramaprasad, Royce Sadler, Michael Scriven, and Dylan Wiliam. The purpose of this chapter is to provide a thorough review of the pertinent literature related to the formative use of summative assessments from the instructor's perspective. The information is organized into the following sections: (1) background of assessment, (2) formative use of summative assessment, (3) using feedback to adjust teaching, and (4) using grades to change student motivation.

Background of Assessment

Formative assessment has an extensive research base that draws on both cognitive and motivational research (Shepard, 2005) but few studies have been conducted with students in higher education (Yorke, 2008). Although there is a growing base of assessment literature that instructors can consult, most of it is at the primary and secondary levels in the U.S. and Britain (Martin & Collins, 2011).

History of Assessment

The history of assessment is grounded in the history of evaluation. Guba and Lincoln (1989) outlined a four step history of evaluation. First generation evaluation marked the period up until World War I. It is described as the era of measurement, where students were characterized as objects. Tests were used to ascertain the students' content mastery. Shortly after World War I, the second generation of evaluation began, the era of description. Second generation evaluation techniques were objective-oriented. Early in the post-Sputnik period, third generation evaluation, with its emphasis on judgment and the standards upon which judgments were made, was born. The first three evaluation generations were described as being based in the modernist tradition of closed systems with an emphasis on control. Fourth generation evaluation

was based on a post-modern, constructivist paradigm typified by open systems with an emphasis on empowerment.

Michael Scriven (1967) first suggested the distinction between formative and summative approaches in reference to evaluations of curriculum and teaching methods. He observed that while evaluation “may have a role in the on-going improvement of the curriculum” (p. 41), it could also be used to examine the curriculum of an entire program. Based on this recognition, he introduced the terms ‘formative evaluation’ and ‘summative evaluation’ to differentiate these distinct roles (Scriven, 1967, p. 43). He suggested that evaluators could gather information early in the process of implementation to identify areas for improvement and adaptation, and at successive stages of development. In a review of the formative assessment literature from French-speaking countries, Allal and Lopez (2005) traced the history of formative assessment from Scriven’s (1967) original definition of “formative evaluation” of educational programs, noting that the term “assessment” had “progressively replaced ‘evaluation’ when the object is student learning in the classroom” (p. 241).

Benjamin Bloom (1969) is credited with suggesting that the distinctions Scriven applied to curriculum could also be applied to the evaluation of student learning, or what is commonly referred to today in the United States as ‘assessment’ (Allal & Lopez, 2005; Knight, 1995; Wiliam, 2006, p.283). Bloom (1969) and Bloom, Hasting and Madaus (1971) adopted Scriven’s idea, applying the concept to student assessment in their work on “mastery learning.” They initially proposed that instruction be broken down into successive phases and students be given a formative assessment at the end of each of these phases. Teachers would then use the assessment results to provide feedback to students on gaps between their performance and the “mastery” level, and to adjust their own teaching to better meet identified learning needs (Allal, 2005).

Newton (2007, p. 152) detailed that in their publication, *Handbook on Formative and Summative Evaluation of Student Learning*, Bloom et al. (1971) identified three characteristics to distinguish between formative and summative:

1. *purpose* – expected uses to which the outcomes will be put (formative assessment focuses on helping the learner learn while summative assessment focuses on grading or certification)

2. portion of course covered – *timing* (formative assessment tends to be more frequent, focusing on smaller units of instruction and occurring during a course rather than at the end)
3. level of *generalization* sought by items in the instrument used to collect data for the assessment (formative assessment focuses on testing for narrow components of proficiency while summative assessment focuses on testing for broad abilities)

Towards the end of the 1980s, Royce Sadler developed a theory of formative assessment based on the conditions for effective feedback (Newton, 2007). He chose not to emphasize timing as did Bloom et al. (1971) and Scriven (1967) using purpose and effect to distinguish between formative and summative assessment (Sadler, 1989):

Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial-and-error learning. Summative contrasts with formative assessment in that it is concerned with summing up or summarizing the achievement status of a student, and is geared towards reporting at the end of a course of study especially for purposes of certification. It is essentially passive and does not normally have an immediate impact on learning, although it often influences decisions which may have profound educational and personal consequences for the student. The primary distinction between formative and summative assessment relates to purpose and effect, not to timing (p. 120).

Learning Theories and Assessment

James and Lewis (2012) explained that learning theories can be grouped into three views, each with different implications for assessment. In literature from the United States (Bredo, 1997; Greeno, et al., 1996; Pellegrino et al., 2001) these three perspectives are generally categorized as 'behaviorist,' 'cognitive' and 'situated.' In literature from the United Kingdom, the same groupings are labeled 'behaviourist,' 'constructivist' and 'socio-cultural.' Watkins (2003) described these three views of learning as: (1) Learning is being taught; (2) Learning is individual sense-making; and (3) Learning is building knowledge as part of doing things with

others. James and Lewis (2012) suggested that “teachers search for assessment models and practices to support their educational goals and the processes of learning they value” (p. 202).

Behaviorism, or what Watkins (2003) described as ‘learning is being taught’, views learning as the conditioned response to external stimuli, using reward and punishment to form or extinguish habits. Using this approach, learning is best accomplished when complex wholes are deconstructed and each part studied separately and then built upon progressively. Behaviorism is often equated with the memorization of information or behavior change in progressive levels of skills. The use of practice tests to increase scores on national tests follows behaviorist assumptions about learning (James & Lewis, 2012). Formal testing relies heavily on behaviorist approaches with poor performance often being remediated by more practice on incorrect answers by further deconstruction and a return to basic skills. Whereas behaviorism has been falling out of favor as a learning theory, cognitive theories currently have a much larger group of advocates (James & Lewis, 2012).

Cognitive constructivist theories or ‘learning is individual sense-making’ (Watkins, 2003) focused on how people construct meaning and make sense of the world by developing mental models. Experts organize relevant knowledge into mental models that make it easier to retrieve and more useful and problem solving is regarded as the context for knowledge construction. Prior knowledge is seen as a powerful determinant of a student’s capacity to learn new material. James and Lewis (2012) argued that it is the importance of prior learning that makes formative assessment an integral part of cognitive learning theory because it is necessary to draw out student’s mental models “in order to scaffold their understanding of knowledge structures and to provide them with opportunities to apply concepts and strategies in novel situations” (p. 191). Teaching and formative assessment works together in this way to close gaps between current understandings and sought after new understandings.

In the situated or socio-cultural perspective, where ‘learning is building knowledge as part of doing things with others’ (Watkins, 2003), learning occurs in interactions between individuals and the social environment. James and Lewis (2012) maintained that the origins of this perspective can be traced back to John Dewey and philosophical pragmatism. Dewey’s work influenced Vygotsky (1978), who contended that since language is developed in relationships between people then social relationships precede and are necessary for learning to occur. Therefore, learning is social and involves participation, and is a collaborative activity in which

people develop their thinking together (James and Lewis, 2012). In this perspective, learning is a mediated activity in which cultural artifacts have a prominent role (Wenger, 1998). James (2008), who originated the term ‘third generation assessment’, suggested the following guidelines to implement situated classroom assessment:

- If learning cannot be separated from the actions in which it is embodied, then assessment too must be ‘situated.’
- Assessment alongside learning implies that it needs to be done by the community rather than by external assessors.
- Assessment of group learning is as important as the learning of the individual.
- ‘In vivo’ [in a natural setting] studies of complex problem solving may be the most appropriate form for assessments to take.
- The focus should be on how well people exercise ‘agency’ in their use of the resources or tools to formulate problems, work productively and evaluate their efforts.
- Learning outcomes can be captured and reported through various forms or recording, including narrative accounts and audio and visual media. The portfolio has an important role here.
- Evaluation needs to be more holistic and qualitative, not atomized and quantified as in measurement approaches. (James, 2008, p.31)

The Assessment Process

Assessment in education is commonly defined as an ongoing process. In the spring of 1995, Thomas Angelo, then director of the American Association for Higher Education (AAHE) Assessment Forum presented a draft definition of assessment to colleagues and solicited their comments. Based on the feedback he received, Angelo published the following definition in the AAHE Bulletin (1995) that is still widely used in higher education today (Suskie, 2009):

Assessment is an ongoing process aimed at understanding and improving student learning. It involves making our expectations explicit and public; setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and

standards; and using the resulting information to document, explain, and improve performance (Angelo, 1995, p.7).

Suskie (2009) adapted the AAHE's definition and developed figure 2.1 below which represents assessment as a continuous four-step cycle. In step four, assessment results are used to review and if needed, revise approaches to the other three steps (p.4):

Figure 2.1 Teaching, Learning, and Assessment as a Continuous Four-Step Cycle

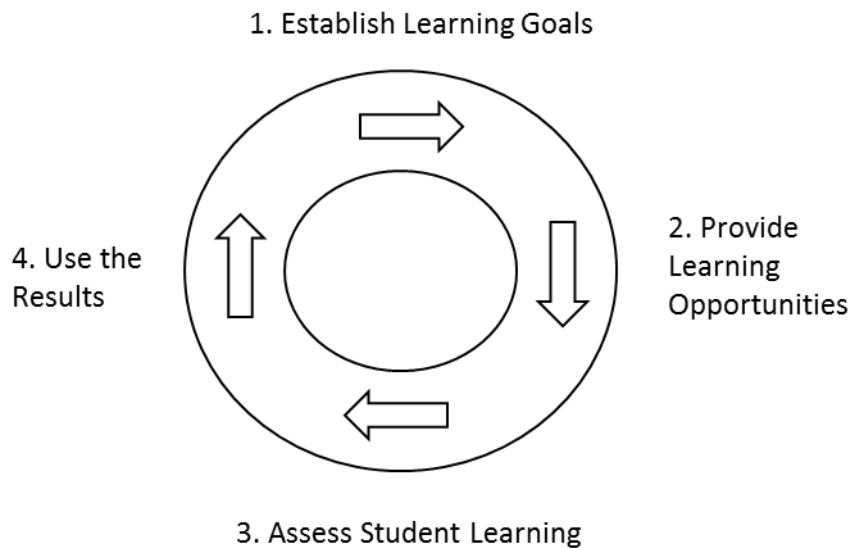


Figure 2.1. The continuous four-step cycle of teaching, learning, and assessment. Source: Suskie, L. (2009, p. 4). *Assessing Student Learning: a common sense guide*, (2nd ed.). San Francisco: Jossey-Bass.

Burke (2010) agreed with Suskie, defining assessment as an ongoing “process of gathering evidence of student learning to inform instructional decisions” (p. 19). Assessment consists of all the tools that teachers use to collect information about student learning and instructional effectiveness. Tests, presentations, observations, and classwork are used to assess student learning. The instructional purpose of a formative assessment is to provide feedback during the learning process; the instructional purpose of a summative assessment is to make a final judgment at the end of the learning process.

Drawing on Scriven (1967), Taras (2005) argued that the process of assessment is the same process used in both summative and formative assessment. Scriven (1967) described assessment as a single process, “Evaluation is itself a methodological activity which is essentially similar whether we are trying to evaluate coffee machines or teaching machines, plans for a house or plans for curriculum” (p.40). Taras (2005) concurred, and asserted that although Scriven was speaking in the context of curriculum evaluation, the principles he discussed are “universally relevant to all assessment” (p. 468).

Assessment was initially thought of as separate from the teaching and learning process with tests or examinations coming at the end of a study unit (Looney, 2011). Initial work on formative assessment in the early 1980s changed this approach somewhat by incorporating tests within study units. Although this made it possible for teachers to diagnose learning needs and adjust teaching, until recently assessments have been seen as being separate from normal classroom activities (Looney, 2011).

Research in higher education has pointed out problems with the assessment process. Richard James (2003), who conducted a major survey of higher education assessment practices in Australia, concluded that, “assessment is one of the least sophisticated aspects of university teaching and learning” (p. 197). The Quality Assurance Agency for Higher Education in the United Kingdom has consistently shown that assessment is the aspect of curriculum that stands most in need of development (Yorke, 2008).

Contemporary Assessment

Although assessment is of central importance in education (Taras, 2005), there is a lack of commonality in the related definitions and terminology. In primary, secondary, and higher education, assessment has many meanings that are contextual in nature. Assessment is an ongoing process that can occur before, during, and/or after instruction is delivered. Assessment can focus on individual student learning, on an entire class of students, a program, the educational institution, or an educational system as a whole. The final purposes and assessment practices in education depends on the theoretical framework of the practitioners and researchers, their assumptions and beliefs about the nature of the human mind, the origin of knowledge and the process of learning.

“Because the assessment of student learning in higher education is relatively new compared to many other fields of study, and because it has been undertaken by people from disciplines with widely differing orientations, the vocabulary of assessment is not yet standardized” (Suskie, 2009, p. 3). Gardner (2012) noted that several different descriptions for assessments are used based on purposes, uses, and functions. Harlen (2007) defined assessment in terms of purpose being the reason for the assessment and use being what is actually done with the results. Wiliam and Black (1996) choose to refer to the functions of assessment while Black et al. (2002) defined assessment in terms of the use made of the evidence. Newton (2010) combined definitions and didn’t attempt to distinguish between uses and purposes. Taras (2010), noted that the purpose can be decided prior to, during or even after the assessment.

Assessments are most commonly classified as summative or formative, based on the function or purpose they will serve. Summative assessments are graded events that occur at chosen points in instruction to determine what students have learned. Formative assessment is part of the instructional process and includes feedback. Balanced assessment is combining the use of both summative and formative assessment practices to gain a better understanding of student learning. Diagnostic assessments are used to determine a student’s knowledge prior to instruction.

Marzano (2010) further classified assessments into three types: obtrusive, unobtrusive, and student-centered. Obtrusive assessments interrupt the flow of activity in the classroom. Instruction stops and does not occur during while students “take the assessment” (p. 23). In contrast, unobtrusive assessments do not interrupt the flow of activity in the classroom. For example; a teacher observing a student stop a soccer ball with her feet and then kick it to a teammate during a game would provide the opportunity for an unobtrusive assessment. The third type, which Marzano considered the most underutilized, are student-generated assessments. As the name implies, students generate ideas about the manner in which they will demonstrate their current status on a given topic. For example; a student might propose that she design and explain a leadership model to demonstrate her knowledge of the topic.

The terms summative and formative assessment are now fundamental to understanding assessment in education. Summative assessment focuses on summing up or summarizing student achievement (Bloom, Hastings, & Madus, 1971; Sadler, 1989; Shavelson, 2006), while formative assessment is centered on active feedback that assists learning (Black & Wiliam, 2004;

Sadler, 1989; Shavelson, 2006). Many educators refer to summative assessment as assessment *of* learning and formative assessment as assessment *for* learning (Black & Wiliam, 2003; Broadfoot, 2008; Stiggins, 2002).

Summative Assessment

Harlen (2004b), felt the term ‘summative assessment’ referred to an assessment with a particular purpose – that of providing a record of a student’s overall achievement in a specific area of learning at a certain time. “It is the purpose that distinguishes it from assessment described as formative, diagnostic, or evaluative. Thus a particular method for obtaining information, such as observation by teachers, could, in theory, be used for any of these purposes and so does not identify the assessment as formative or summative” (p.12).

Taras (2005) argued that all assessment begins with summative assessment and that formative assessment is summative assessment plus feedback which is used by the learner. “The process of assessment leads to summative assessment, that is, a judgment which encapsulates all the evidence up to a given point. This point is seen as finality at the point of the judgment” (Taras, 2005, p.468). In a survey of instructors in the Education department at an English university, Taras (2011) found that: There was a general agreement of summative assessment: 80 % (40/70) mentioned the word ‘end’ or ‘final’, 36 % (18/47) mentioned ‘grade’ (p.176).

Ainsworth and Viegut (2006) asserted that classroom summative assessments (given by individual instructors) and common summative assessments (developed by teaching teams) take place at the end of a learning period after all instruction is presented and therefore by design and purpose are not intended to improve current student learning:

If all instruction and related learning activities for the particular standards have concluded, the results of summative assessments are not used to improve student understanding for *current* students. Instead, teachers typically use these assessment results to judge the effectiveness of their teaching practices and to improve instruction of those standards for *future* students (emphasis from original) (p.24).

Harlen (2012) proposed a framework to conceptually represent an assessment for summative purposes noting that in practice, distinctions between summative and formative assessment are not so clear. Based on Harlen (2006), figure 2.2 illustrates that the primary use of

a summative assessment is to report on what has been achieved rather than inform teaching or learning. Evidence related to learning goals is gathered from tests, tasks, or regular activities, and interpreted in terms of a judgment of achievement, using the same criteria for all students (criterion-referenced). The judgment or interpretation may be by instructors, or external agencies, with no immediate feeding back into teaching, and students have no role in the assessment. Unlike the formative assessment model which is a closed loop (Fig 2.3), the summative assessment model is open (Harlen, 2012, p. 91):

Figure 2.2 Assessment for Summative Purposes

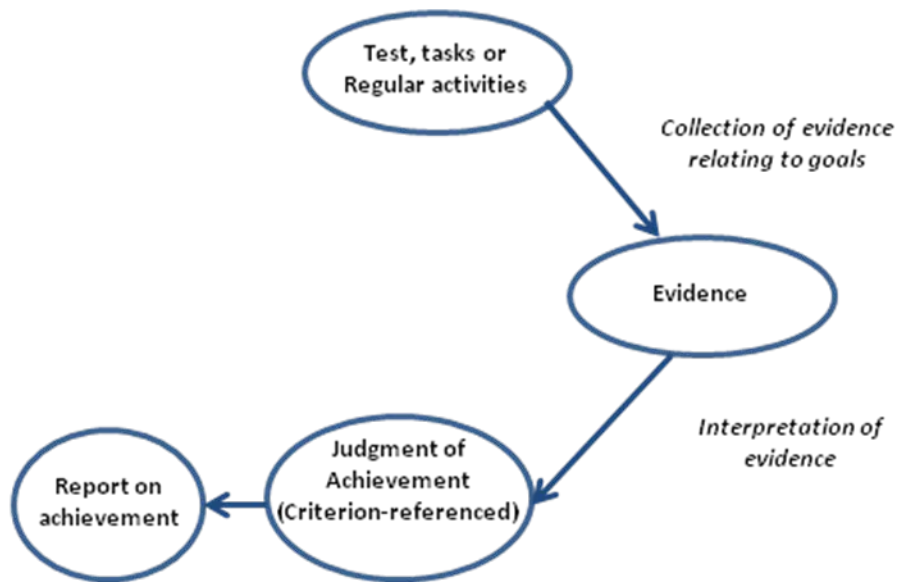


Figure 2.2. Source: Harlen, W. (2012, p.91). On relationships between assessment for formative and summative purposes, in Gardner, J. (ed). *Assessment and learning* (2nd ed.). Los Angeles: Sage.

Formative Assessment

Formative assessment has many definitions:

1. "An assessment activity can help learning if it provides information to be used as feedback by teachers and by their pupils in assessing themselves and each other to modify the teaching and learning activities in which they are engaged. Such

- assessment becomes 'formative assessment' when the evidence is actually used to adapt teaching work to meet learning needs" (Black, Harrison, Lee, Marshall & Wiliam, 2003, p.2).
2. "The process used by teachers and students to recognize and respond to student learning in order to enhance that learning during the learning" (Bell & Cowie, 2000, p.536).
 3. "Formative assessment is defined as assessment carried out during the instructional process for the purpose of improving teaching or learning" (Darling-Hammond & Rust, 2005, p.275).
 4. "Formative assessment is concerned with how judgments about the quality of student responses (performances, pieces, or works) can be used to shape and improve the student's competence by short-circuiting the randomness and inefficiency of trial and error learning" (Sadler, 1989, p.120).

The landmark review by Paul Black and Dylan Wiliam (1998a) is the most widely cited reference on formative assessment (Shepard, 2005). Black and Wiliam examined the findings from more than 250 studies from research literature addressing classroom practices; student motivation and participation in assessment practices; learning theory; and the properties of effective feedback. They concluded that formative assessment has a more profound effect on learning than do other typical educational interventions:

The research reported here shows conclusively that formative assessment does improve learning. The gains in achievement appear to be quite considerable, and as noted earlier, among the largest ever reported for educational interventions. As an illustration of just how big these gains are, an effect size of 0.7, if it could be achieved on a nationwide scale, would be equivalent to raising the mathematics attainment score of an "average" country like England, New Zealand, or the United States into the "top five" after the Pacific rim countries of Singapore, Korea, Japan, and Hong Kong (Black & Wiliam, 1998a, p. 61).

Harlen (2012, p. 90) used Figure 2.3, based on Harlen (2006), to show how assessment for formative purposes is a cycle of events. Evidence is gathered in activity A and then interpreted in terms of progress towards lesson goals. Some notion of progression in relation to

the goal is needed for this interpretation, so that where students have reached can be used to indicate what next step is appropriate. Helping students to take this next step, leading to activity B, shows how the evidence of current learning is fed back into teaching and learning. This feedback helps to regulate teaching so that the pace of moving towards a learning goal (step C) is adjusted to ensure the active participation of the students. As with all regulated processes, feedback into the system is the important mechanism for ensuring effective operation.

Figure 2.3 Assessment for formative purposes

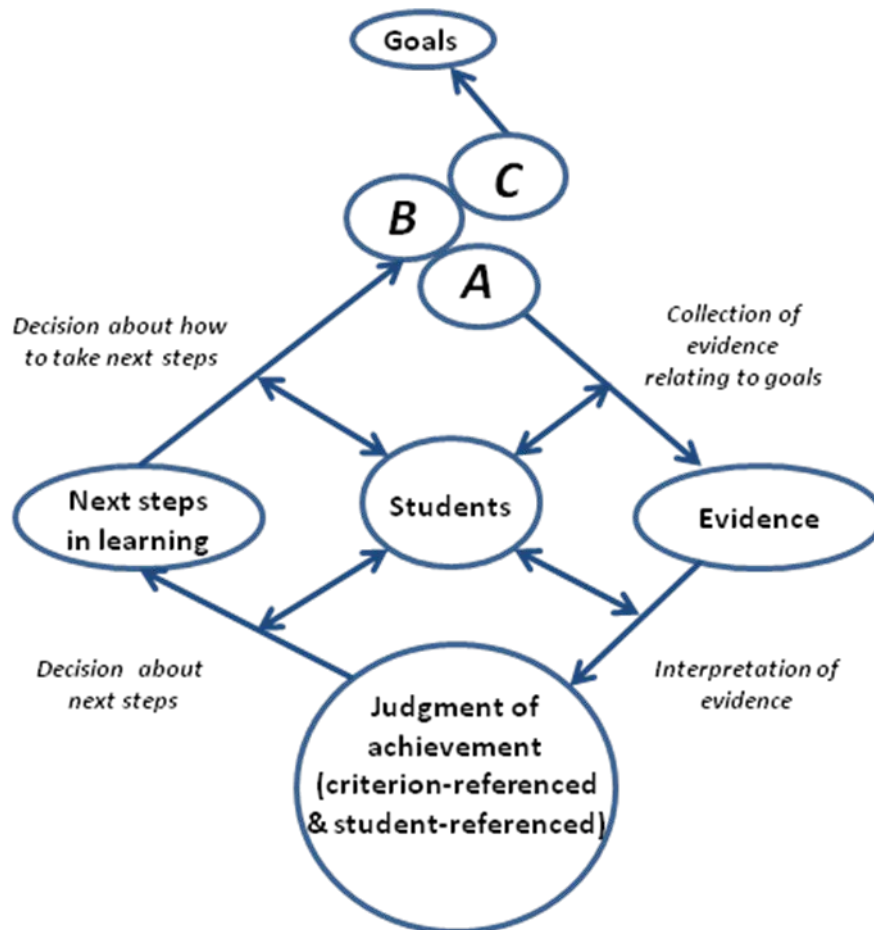


Figure 2.3. Source: Harlen, W. (2012, p. 90). On relationships between assessment for formative and summative purposes, in Gardner, J. (ed). *Assessment and learning* (2nd ed.). Los Angeles: Sage.

Dunn and Mulvenon (2009) challenged the results of Black and William's (1998a) review, explaining that while their research did provide some support for the positive impact formative assessment had on student achievement, it mostly pointed out the need for more research using more efficient methodologies. They argued not that formative assessment is unimportant, but "that limited empirical evidence exists to support the 'best practices' for formative [assessment]" (Dunn and Mulvenon, 2009, p. 9).

Other studies have concluded that effective implementation of formative assessment may be more the exception than the rule (Black, 1993; Black & Wiliam, 1998b; Stiggins et al., 1989). The quality of formative assessment rests, in part, on strategies teachers use to elicit evidence of student learning related to goals, with the appropriate level of detail to shape subsequent instruction (Bell & Cowie, 2001; Heritage, 2010; Herman et al., 2010). It is much more typical to find that teachers emphasized rote learning, develop only superficial questions to probe student learning, and provide only general feedback. Teachers may have difficulty in interpreting student responses or in formulating next steps for instruction (Herman et al., 2010). While many teachers agree that formative assessment methods are an important element in high quality teaching, they may also protest that there are too many logistical barriers to making formative assessment a regular part of their teaching practice, such as large classes, extensive curriculum requirements, and the difficulty of meeting diverse and challenging student needs.

Black & Wiliam (2009) put forth the theory of formative assessment below explaining that its focus is very narrow because, "...it helps distinguish a theory of formative assessment from an overall theory of teaching and learning" (p.8):

Practice in a classroom is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they would have taken in the absence of the evidence that was elicited (Black & Wiliam, 2009, p.23).

Assessment and Learning

Bloxham and Boyd (2007) asserted that understanding the link between assessment method and student approach to learning is essential for the effective design of assessment in higher education. When assessment is the most influential element of the learning environment

and the learning context is altered (Elton & Johnston, 2002), it is likely that students' approach to learning will change (Prosser & Trigwell, 1999). Biggs (2003) offered that an individual's approach to learning is not a fixed characteristic, but is influenced by the perception of the given assessment task.

Bloxham and Boyd (2007) explained that the two approaches to learning used by students and most often discussed in the associated literature are a surface approach and a deep approach (Marton & Saljo, 1997; Ramsden, 2003). A student's notion of learning and their intention determine which approach to learning will be used and when. Students use a surface approach when their intention is to address work requirements with little desire to understand the material. Students who use this approach want to succeed with the minimum amount of effort. These students tend to focus on the knowledge level and memorize information for use during an upcoming exam. The result is a limited understanding of the material and a failure to grasp the overall meaning of their studies (Entwistle, 1997). In contrast, students who use a deep approach to learning desire to understand ideas and are inherently interested in their studies. These students relate information and ideas to their own experiences and look for patterns, principles and meanings resulting in higher-quality learning outcomes. Properly designed assessment can encourage students to use a deep approach to learning while the opposite is true for poorly designed assessment (Bloxham & Boyd, 2007).

Table 2.1 Compilation of Assessment Characteristics

	Summative Assessment	Formative Assessment
Purpose	Evaluate student learning for accountability, grading, and certification.	Check student progress to improve learning.
	Compare against a standard or benchmark with points/grades.	Interactive, focus is on mastery of classroom content.
	Reports on what has been achieved, no immediate effect on learning.	Observation. Catches deficiencies. Identifies a gap in what students need to know and know.
Timing	“Sums up” learning at a certain point in time.	While program activities are “forming.”
	End of learning period, comprehensive.	More frequent, focus on smaller portions of content. During the course of learning.
Level of generalization	Testing for broad abilities.	Narrow components of proficiency.
Responsiveness	Too late to affect learning.	Responsive to student needs. Instructors can adjust instruction activities.
Grades	Often results in a grade with high percentage overall course value.	If graded, usually low percentage overall course value.
Feedback to student	Grade in Ten working days or less is CGSOC standard. CGSC 1002 Assessment Form usually includes rubric.	Must have feedback to be considered formative. Descriptive. Oral or written. Immediate, in classroom.
Feedback to instructor	Used to plan future instruction.	Used to plan next instruction.
Motivation	Extrinsic, performance oriented... grades.	Intrinsic, mastery-oriented.
Learning	Summarizes learning.	Helps learning.
Examples	Mid-term or final exam.	Journal, Think-pair-share, write-pair-share, reflection.
Assessment instrument	Formal. Common to all. Valid and reliable. Created prior to instruction.	Informal, can be individually produced.
Also known as	Assessment of Learning, High risk.	Assessment <i>for</i> Learning, Low risk.
Results used	Externally and internally. Are leaning goals being met?	Internally to improve teaching and learning.

Table 2.1. Source: Compilation of summative and formative assessment characteristics from the literature review. Some characteristics conflict as is the case with the literature.

Evolving from evaluation, contemporary assessments were commonly categorized as either summative (of student learning) or formative (for student learning). Because of the contextual nature of assessment and a lack of commonality in terms, there is no one agreed upon definition for educators in primary or higher education to use to inform their communities of practice. Proponents of the formative use of summative assessment recognized that the same assessment could be used for both summative and formative purposes.

Formative Use of Summative Assessment

Paul Black (1998), who has contributed a great deal to debates on formative and summative assessment, described the distinction between formative and summative assessment as being one of purpose/function (Newton, 2007). Black (1998) argued that the two functions are at the ends of the same spectrum:

Some have laid stress on the differences between the formative and summative purposes, and have argued that the assessment instruments and procedures needed for one are so different from those needed for the other that neither can flourish without clear separation. On the other side, it can be argued that the two functions are at the ends of the same spectrum and that there is no sharp difference, and that if the two functions are separated, then teachers' assessment work will be devalued (p. 34).

Harlen and James (1997) distinguished between the characteristics of formative and summative and clarified the relationship between the two types of assessment. They asserted that while it is not possible to aggregate assessment judgments made for formative purposes to derive summative judgments, the same evidence can be used to derive formative and summative judgments provided the evidence is examined separately by purpose (Newton, 2007). Harlen (2005) later developed and further clarified the distinction between formative and summative:

The two main purposes of assessment discussed in this article are for helping learning and for summarizing learning. It is sometimes difficult to avoid referring to these as if they were different forms or types of assessment. They are not. They are discussed separately only because they have different purposes; indeed the same information, gathered in the same way, would be called formative if it were used to help learning and teaching, or summative if it were not so utilized but only employed for recording and

reporting. While there is a single clear use if assessment is to serve a formative purpose, in the case of summative assessment there are various ways in which the information about student achievement at a certain time is used (p. 208).

Many educators assert that summative assessment happens too far along in the learning process to make instructional adjustments (Ainsworth & Viegut, 2006; Burke, 2010; Garrison & Ehringhaus, 2011). Ainsworth and Viegut (2006) argued that since all the learning activities related to the targeted standards have concluded, “the results of summative assessments are not used to improve student understanding for current students. Instead, teachers typically use these assessment results to judge the effectiveness of their teaching practices and to improve instruction of those standards for future students” (p. 24). Alternatively, Knight (2000) believed that tasking students to do an assignment for formative feedback only, will not motivate them to work hard. He offered that including a “mark” on the student’s work will make it both summative and formative. Careless (2006) described pre-emptive formative assessment in which it is primarily the teachers who learn from the evidence and adjust their teaching.

However, research does not reflect the attitude that summative assessment happens too far along in the learning process to allow for instructional adjustments. Black et al. (2003) adjusted their ideas on using formative and summative assessment to support each other in the classroom context because teachers refused to separate them (Black et al., 2004; Taras, 2005) and research evidence showed that teachers found it useful to use summative assessment for formative purposes (Black et al., 2004, Taras, 2001).

An increasing volume of literature advocates use of the same assessments for both summative and formative purposes (Burke, 2010; Stiggins, 2009; Taras, 2008). Drawing on Harlen (1998), Harlen (2012) used Figure 2.4 to illustrate how the “blurred distinction between assessment to help learning and assessment to report learning indicates that the relationship between formative and summative assessment might be better described as a ‘dimension’ rather than a ‘dichotomy’” (p. 98).

Figure 2.4 A possible dimension of assessment purposes and practices

Formative ←-----→Summative				
	Informal formative	Formal formative	Informal summative	Formal summative
Major focus	What are the steps in learning?		What has been achieved to date?	
Purpose	To inform next steps in learning	To inform next steps in teaching	To monitor progress against plans	To record achievements of individuals
How evidence is collected	As normal part of class work	Introduced into normal class work	Introduced into normal class work	Separate task or test
Basis of judgment	Student- and criterion-referenced	Student- and criterion-referenced	Criterion-referenced	Criterion-referenced
Judged by	Student and teacher	Teacher	Teacher	Teacher and external marker
Action taken	Feedback to students and teacher	Feedback into teaching plans	Feedback into teaching plans	Report to student, parent, other teachers, etc.
Epithet	Assessment for learning	Matching	Dip stick	Assessment of learning

Figure 2.4. Source: Harlen, W. (2012, p. 98). On relationships between assessment for formative and summative purposes, in Gardner, J. (ed). *Assessment and learning* (2nd ed.). Los Angeles: Sage.

Harlen (2012) explained that the outside columns of Figure 2.4 show practices and uses that are typically associated with assessment for learning and assessment of learning. Between these two extremes, in the formal formative and informal summative columns, there is a range of activities that may be similar in practice, but have varying roles in teaching and learning. Because there are varying degrees of formality, what is described as informal summative may involve the same instructor activity that is used to describe formal summative (see Fig 2.4). The key difference is the use of evidence collected. If the evidence collected, for example, a classroom test is used to adapt teaching, the cycle is closed (see Fig 2.3) and the assessment is formal formative. If evidence collected from the same test results in no feedback into teaching, (see Fig 2.2) than the assessment is categorized as informal summative. Harlen (2012) noted

similarities between this approach and Cowie and Bell's (1999) observations which indicated that formative assessment can be classified as either planned or interactive. Similar to formal formative, a planned formative assessment is used by the instructor for information gathering purposes and the results are fed back into teaching. Interactive formative assessments and informal formative assessments are both unplanned and can be immediate, with feedback going to both student and instructor.

Harlen (2012) clarified that recognition of how evidence collected can be used for both summative and formative purposes should not be seen as a reason against retaining the current distinctions between these types of assessment. Harlen (2012) argued that:

In both cases there are limitations in the dual use of the evidence, but of rather different kinds. The limitation of using evidence which has initially been gathered for a summative purpose to help learning bears on the validity of the evidence; it is just not sufficiently rich and readily available to be adequate for formative use. The limitation of using evidence which has initially been gathered by teachers to help learning to report on learning bears on the reliability of the evidence (p. 99).

After examining textbooks on program planning in adult education by Caffarella (2002) and Gailbraith et al. (2001), Martin and Collins (2011) concluded that the authors described formative evaluation of a course or program as being either "in progress" or "ongoing," referring to the interval of time from start to finish. Contrasting this to Wiliam's (2006) model of formative assessment, created for secondary education, Martin and Collins detailed how Wiliam used both length and focus to identify three types of formative assessment: long-cycle, medium-cycle, and short-cycle. Martin and Collins asserted that Wiliam's construct can be applied to adult education programs, including graduate programs. Their adaptation of Wiliam's (2006) Types of Formative Assessment Model is Figure 2.5 below.

Figure 2.5 Accelerated Adult Education Version of Types of Formative Assessment

Type	Focus	Length
Long-cycle or more	Across courses, programs, years	8 weeks to 1 year
Medium-cycle	Within and between courses	1 to 8 weeks
Short-cycle	Within and between classes	5 sec to 1 week

Figure 2.5. Source: Martin, J., & Collins, R.A. (2011, p. 132). Assessing and evaluating adult learning in career and technical education, in Wang, V.X. (ed.). *Formative and summative evaluation in the assessment of adult learning*. Hershey, PA: IGI Global.

William’s (2006) model, like Martin and Collin’s adaptation (Figure 2.5), allows for formative assessment to occur in the middle of a class period, and it allows for information gathered from what is usually called summative assessment to be used as formative assessment if it leads to actions that change or improve learning. Martin and Collins (2011) asserted that this highlights a key difference between those who use time as the determining factor between formative and summative assessment and those who use the purpose of the assessment as the determining factor. Martin and Collins (2011) further explained that timing should not be the determining factor in identifying an assessment as formative or summative:

A teacher could give an examination at the end of the semester which is for the purpose of assigning grades (summative) and then use the data gathered from that examination to improve her curriculum for the next semester and turn the summative assessment into a formative assessment. Time is not an issue for her, as she is utilizing the data derived from her summative assessment to create change and improvement, which now meets our definition of a formative assessment (p. 132).

Ainsworth and Viegut (2006) noted that educators are often confused when trying to classify an assessment given at the end of a learning period. They offered that commonly a broad distinction based on purpose is used: “If the *results* from that assessment can be used to monitor and adjust instruction in order to improve learning for *current* students, the assessment can be

said to be formative. If not, the assessment is summative” (Ainsworth & Viegut, p. 27). They contended that classification of an assessment depends on not only the assessment’s purpose but also on “*how it is to be used.*” They offer three examples to demonstrate this distinction:

1. If the assessment is simply a final measure of how students performed on multiple standards taught during the quarter, semester, or trimester course of study, the assessment is obviously summative.
2. If a teacher uses the results from a unit test in any way to inform instruction for the same students during the next unit of study, the test results are being used formatively, even though the test itself was a summative measure used to determine student understanding of the particular content taught during that unit.
3. If a teacher provides students with the opportunity to revise and thus improve their performance on a particular assessment during the evaluation process, the assessment can rightly be considered formative. Once the students complete their revisions and the final evaluation is determined, the assessment is now summative (Ainsworth and Viegut, 2006, p. 27).

Harlen (2012) advocated that the principles for learning developed by the Assessment Reform Group (ARG, 2002) can be used to check the extent to which evidence from a summative assessment can be truly formative and used as such. Positive answers to the questions below will enable an instructor to judge a summative assessments’ ability to help student learning or assessment *for* learning (p. 94):

- Does it focus on how students learn?
- Is it sensitive and constructive?
- Does it foster motivation?
- Does it promote understanding of goals and criteria?
- Does it help learners to know how to improve?
- Does it develop the capacity for self-assessment?
- Does it recognize all educational achievements?

Assessment shapes the experience of students and influences their behavior more than the teaching they receive (Bloxham and Boyd, 2007). For this reason “there is more leverage to improve teaching through changing assessment than there is in changing anything else” (Gibbs and Simpson, 2004, p. 22). The use of graded or summative assessment for formative purposes is an area of assessment often described as underused but with much potential for improving student learning (Black et al., 2004; Stiggins, 2009; Taras, 2008).

Balanced Assessment Model

Focused at the primary and secondary levels of education, Kay Burke (2010) and Stiggins et al. (2004) advocated the use of a balanced assessment model. In a balanced assessment model the same assessment could be classified as formative during the initial teaching and learning process and as summative during the final stages of the grading period. Formative and summative assessments support each other as Burke (2010) explained “they can even be the exact same thing - only the purpose and the timing of the assessment determine its label” (p. 24). This was found to be especially important in primary and secondary education where students who practice on assessments that are very similar to standardized state or district assessments tend to score higher (Harlen, 2003).

According to Stiggins et al. (2004) “a balanced assessment system takes advantage of assessment of learning and assessment for learning: each can make essential contributions. When both are present in the system, assessment becomes more than just an index of success. It also serves as the cause of that success” (p. 25). The direct benefit of a balanced assessment system is that the integration of both formative and summative assessments allows teachers to use feedback from formative activities to modify their instruction to help students achieve the standard (Burke, 2010).

Common Assessments

Common assessments are formative or summative assessments that are created collaboratively by teams of teachers who teach the same class or grade level. They are most often used as school-level assessments to provide evidence to lead teachers, curriculum personnel, and administrators that desired standards are being met (Burke, 2010; Stiggins and DuFour, 2009). Common assessments can be formative or summative, and the results can be used by teachers, curriculum personnel, or administrators as evidence that standards are being met. When focused on important standards that are aligned with summative or high-stakes assessments, common formative assessments can inform instruction and predict student results (Burke, 2010).

Ainsworth and Viegut (2006) asserted that once teachers realize the positive impact on student learning resulting from the use of formative assessment, they will find the time to integrate and continue this “powerful practice” (p. 3). Stiggins and DuFour (2009) argued that

common assessments for formative purposes, or “assessments created collaboratively by teams of teachers who teach the same course” (p.640) are a powerful tool for creating effective assessments.

Embedded and Add-on Assessments

Suskie (2009) detailed the use of embedded and add-on assessments. Usually designed by local faculty and staff, embedded assessments are program, general education, or institutional assessments that are embedded into the curriculum. Embedded assessments are designed to do double duty; provide information to teachers on what students have learned in the course and their progress in achieving program or institutional goals (Wilson and Sloan, 2000). Wilson and Sloan (2000), wrote a well-cited study on the Berkeley Evaluation and Research (BEAR) Embedded Assessment System. They explained the BEAR approach to embedded assessments: “By using the term *embedded* we mean that opportunities to assess student progress and performance are integrated into the instructional materials and are virtually indistinguishable from the day-to-day classroom activities” (182).

Add-on assessments are ungraded assessments beyond course requirements that students participate in voluntarily, like portfolios, surveys, or focus groups. Suskie (2009) cautioned that the major drawback with add-on assessments is that students have to be convinced to participate and give the assessment sincere thought and effort. Ekman and Pelletier (2008) advised that while making participation in an add-on assessment a course requirement might make participation appear more important, it will not necessarily compel students to give the assessment their best effort.

Progressive and Continuous Assessment

Maxwell (2004) asserted that progress assessment blurs the boundary between formative and summative assessment. Bell and Cowie (2001) cited progressive or continuous assessment as the approach used by the Queensland, New Zealand Public School System for over 35 years. A school-based system, meaning that outside high-stakes testing is not used.

Under contract to the New Zealand Ministry of Education in 1995-1996, researchers investigated the classroom assessment activities of science education teachers for grades 7-10 (Bell & Cowie, 2001). Researchers found that teachers were unable to explicitly describe what they did in the classroom that was called “formative assessment” (Bell & Cowie, 1997). Cowie

and Bell (1999) proposed that formative assessment should be described as planned or interactive. Teachers either plan assessments before instruction and use them in class or are prepared to exploit situations that arise during interaction with students that arise during instruction.

EPPI-Centre Studies (2002 and 2004)

In 2002 and 2004, The EPPI-Centre (Evidence for Policy and Practice: Information and Co-ordinating Centre), part of the Social Science Research Institute of Education, University of London, conducted the most comprehensive research on the effects of using summative and formative assessments for the same purpose. Led by Wynne Harlen, of the Graduate School of Education, University of Bristol, evidence presented from previous research conducted at primary and secondary schools in English speaking countries was reviewed and three focused studies were published, each discussed in separate paragraphs below (Harlen, 2005, p. 207). The research results of the two 2004 studies were reviewed by a team of educators that included several who are cited often in assessment literature: Paul Black, Kathryn Ecclestone, John Gardner, Lorrie Shepard, Gordon Stobart and Dylan Wiliam.

In the first study, Harlen and Crick (2002) researched 19 studies looking for evidence of the impact of summative assessment and tests on student's motivation for learning resulting in several main findings that are relevant to this study. "First, when preparing students to pass high-stakes tests becomes the focus, teachers tend to use a transmission style of teaching and activities that emphasize knowledge level learning, a disadvantage to students who prefer more active learning experiences. Second, feedback from assessments is important to student learning, but feedback given in a judgmental way may influence student self-efficacy" (Harlen and Crick, 2002, p.2).

These findings are supported by other reviews of research on the effects of high-stakes tests that indicated teachers drift from teaching the curriculum and adopt transmission styles of teaching focusing on training students to do well on a specific test (Black and Wiliam, 1998b; Linn 2000; Stiggins, 1999). Gordon and Rees (1997) maintained that although the results were better on tests students were trained to take; this wasn't proof that students learned more; rather it showed that students can be trained to pass any type of test, even ones that are designed to assess higher levels of thinking (Harlen, 2005).

In the second study, Harlen (2004a) was looking for research evidence of the reliability and validity of assessment by teachers used for summative purposes. After an in-depth review of 30 studies, evidence showed that teachers who participated in developing the criteria on a test were able to use them reliably when evaluating students' work (Hargreaves et al., 1996; Harlen, 2004a) and that the more thoroughly a subject was covered, the more accurately teachers were able to judge student's performance (Coladarci, 1986; Harlen, 2004a). When looking at conditions that affect the reliability and validity of teachers' summative assessment, evidence indicated that moderation (calibration) accomplished through professional collaboration and discussing sources of potential bias benefited reliability (Harlen, 2004a).

Continued interest in giving teachers a greater role in summative assessment prompted the third and final study in 2004. Harlen (2004b) reviewed 23 studies for evidence of the impact the process of using assessment by teachers for summative purposes has on students, teachers and the curriculum. Looking at assessments used for internal purposes, there was evidence that feedback from earlier assessments impacted the amount of effort applied to future tasks of the same kind (Brookhart and DeVoge, 1999; Carter, 1997/8, Harlen, 2004b). Teachers changing assessment processes, practices and explanations they give to students can lead to better student learning (Flex et al., 1995; Harlen, 2004b). Assessment for external purposes has an adverse effect on teachers' performance when it is viewed as too time consuming (Abbott et al., 1994; Bennett and Wragg, 1992).

Findings relative to conditions and contexts communicated the role of the faculty in the classroom and the importance of engaging in collaboration with other faculty members. The process that teachers use in explaining the importance of assessment tasks and in grading impacts student motivation for learning (Brookhart and DeVoge, 1999; Bullock et al., 2002; Harlen, 2004b; Iredale, 1990; Stables, 1992). When embedded into the curriculum, summative assessments have a more positive effect on teachers and their teaching (Bennett et al., 1993; Bullock et al., 2002; Carter, 1997/8; Hall et al., 1997; Harlen, 2004b; Iredale, 1990; Johnston et al., 1993; Koretz et al., 1994; McCallum and McAlister, 1993; Whetton et al., 1991). When teachers are given opportunities to share and develop their understanding of assessment procedures they are able to review and modify teaching practices (Flex and Wragg, 1995; Gipps and Clarke, 1998; Hall et al., 1997; Hall and Harding, 2002; Harlen, 2004b; Heibert and Davinroy, 1993; Valencia and Anu, 1997).

The EPPI-Centre studies of 2002 and 2004 are studies of research of the use of summative assessment and were a valuable resource for this research because of the depth of studies and the long list of assessment researchers who reviewed and commented on the findings. These studies specifically looked at claims made by teachers in the areas of providing formative feedback to students as part of teaching, student motivation, as the dual use formative use of summative assessment.

CGSS Assessment Process

In 2005 the CGSC identified its assessment methodology as a concern in the Self-Study Report submitted to the Higher Learning Commission of the North Central Association of Colleges and Schools for accreditation purposes. The report suggested the need for a more detailed understanding of current faculty assessment processes before CGSC undergoes accreditation again:

As far as the committee could observe, the College focuses enormous effort (much of it misdirected and wasteful) on CGSS core curriculum evaluation and lesson design (student input) while giving only marginal attention to vital aspects of learning assessment (student output)... First an apparent gap exists in student learning outcomes assessment to determine whether program learning objectives are being achieved. Moreover, assessment of student performance by instructors is limited, and some students complained of insufficient feedback from faculty members (U.S. Army CGSC, 2005, p. 182).

Findings from the 2010 Command and General Staff School Faculty Survey provided further evidence of the need to understand how DTAC instructors use standardized assessment instruments to assess student learning. The CGSC Quality Assurance Office invited 307 CGSS faculty members to participate in this survey that consisted of forced-choice, demographic, and multiple open-ended questions designed to capture the faculty's insights and attitudes pertaining to their roles and responsibilities as CGSS faculty members (*CGSS Faculty Survey for AY 2010*, p.1). Of the 178 faculty members that participated in the survey, 32 percent, for a count of 57, were from CTAC (*CGSS Faculty Survey for AY 2010*, p.2), which in 2011 officially changed its name to DTAC. During the period of this research (June 2013), the DTAC had 102 instructors assigned to Ft. Leavenworth.

This survey did not find to the college standard of 66% favorable responses that instructors feel students are; accurately assessed based on the Terminal Learning Objectives /Enabling Learning Objectives; curriculum assessments accurately measure student learning; assessment plans adequately describe how to measure student learning; and assessment plans contain the tools to accurately measure student learning. Of the five CGSS teaching departments surveyed, DTAC was the next to least likely department to find to the college standard. Findings from open ended responses were not classified by teaching department, but respondents did identify courseware/curriculum issues as one of the biggest deterrents to “doing well” in their primary role (*CGSS Faculty Survey for AY 2010*, p.17).

The results of the 2010 Command and General Staff Faculty Survey gave faculty and administration reason to question its grading process. This survey found that instructors did not feel that current assessment instruments accurately measure student learning and current assessment plans do not adequately describe how to measure student learning. This faculty survey did not identify instructor experiences with different types of assessments, explain what instructors thought makes assessments effective, or address the link with feedback provided to students. It was this survey that provided the researcher with the idea for this study.

CGSS Faculty Survey for AY 2010

The *CGSS Faculty Survey for AY 2010* found that survey respondents (178 participated out of 307 invited) did not find to the College standard that:

1. Curriculum assessments they are familiar with are accurate instruments to measure student learning.
2. Assessment plans they are familiar with adequately describe how to measure student learning.
3. Assessment plans they are familiar with contain the tools needed to accurately measure student learning.
4. Students are accurately assessed on their achievement of Terminal and Enabling Learning Objectives (TLO/ELO)

The findings of this survey illustrate the complexity of assessing student learning in the classroom. On the surface the problem appears to be with the assessment instruments or

assessment plans; develop better assessment instruments and plans and the problem is solved. Like most problems in education, this problem is contextual (Suskie, 2009). The common assessments of learning that DTAC instructors use are not all the same type. In the AY 2013 C500 Module there were eight graded assessments. Three of these are based on contribution to group work; three for contributions in practical exercises, one for contributions in a mission analysis and briefing, and one for contribution to a course of action analysis. Of these eight, two were individual assessments requiring scenario analysis, one was a written exam based on a scenario, and one was for completion of an online program designed to impart base knowledge. Each of these assessments required different types and amounts of student preparation, instructor activities and feedback.

The assessment strategy of a particular course has a major impact on student activity (Snyder, 1973). Measuring a desired outcome will do little to improve it without examining the processes that led to the outcome (Banta, Jones & Black, 2009). Bloxham and Boyd (2007) argued that instructors like those at the CGSC who teach approved curriculums with fixed assessments “can have a significant influence on the assessment strategy of the course and how students perceive and engage in assessed tasks” (p. 49). For the purpose of informing practice, this qualitative case study sought to understand, from the instructor’s perspective, the effect using summative assessment has on the activities instructors use to improve student learning.

Using Feedback to Adjust Teaching

Black and Wiliam (1998a) and Gibbs and Simpson (2004-2005) asserted that the inclusion of feedback is the most important aspect of the assessment process when trying to improve student achievement. Research by Hartley et al. (2002) and Weaver (2006) indicated that students value feedback while Gibbs and Simpson (2004-5) contended that students fail to engage with feedback provided. Hounsell (2003) suggested that the combination of students not being able to or choosing not to use feedback, along with faculty cynicism that their efforts are being wasted, thwarts achievement of potential gains from feedback. Bloxham and Boyd (2007) contended that students express considerable dissatisfaction with much of the feedback they receive resulting in little or no impact on their learning.

Wlodkowski (2008) offered that “feedback is probably the most powerful communication that instructors can regularly use to affect learners’ competence” (p. 315). Hattie and Timperley

(2007) concurred and added that most current assessments are not effective because they provide minimal feedback, usually in the form of grades that don't help students understand what they need to do to improve. Although most of the research on feedback has been in the primary and secondary levels (Hattie and Timperley, 2007), it has been shown to be critical for enhancing practice and deepening adult learning (Keeton, Sheckley, & Griggs, 2002).

Paul Black and Dylan Wiliam (1998b) review of over 250 scholarly articles is widely cited as proof that using formative assessment in the classroom improves student learning (Black & Wiliam 1998a; Harlen, 2005). In this same article, Black and Wiliam (1998b) proposed a definition of assessment that followed Ramaprasad (1983) and Sadler's (1989) central use of feedback and connected it to the adaptation of activities by teachers to improve student learning:

We use the general term assessment to refer to all those activities undertaken by teachers – and by their students in assessing themselves – that provide information to be used as feedback to modify teaching and learning activities. Such assessment becomes formative assessment when the evidence is actually used to adapt the teaching to meet student standards. (p. 140)

Feedback can be given formally or informally in a group or a one-on-one setting (Marzano, 2010). Hattie and Timperley (2007) explained that the purpose of feedback is “to reduce discrepancies between current understandings and performance and a goal” (p. 86). Shute (2008) offered that feedback is “information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning” (p. 154).

Race (1993) identified that the greater the quantity of feedback students receive prior to an assessment, the more opportunity students have to learn from that feedback. Race (1995) explained that good feedback is more than just grades; it is discussion and written comments on the graded work. Light and Cox (2001) felt that the scheduling of assessments is usually too late to allow for timely feedback, denying students a sense of their own progression. Feedback that is not timely contributes little to student learning (Chickering & Gamson, 1987).

Assessment is a time-consuming process for all concerned, so it seems like a wasted opportunity if it is not used as a means of letting students know how they are doing, and how they can improve (Brown, Race and Smith, 1996). Chickering and Gamson (1987) promoted the prudent use of model answers, taken either from students' work or specially written by the lecturer, as an excellent form of feedback. Garrison and Ehringhaus (2007) felt that “using

student work, classroom tests, or exemplars of what is expected helps students understand where they are, where they need to be, and an effective process for getting there” (p.2). This technique helps students by engaging them in the learning process by establishing and defining quality work together.

Orsmond and Merry (2011) proposed that students often do not understand instructors’ feedback intentions suggesting that instructors need to provide more guidance on how to effectively use feedback. Although students may revisit feedback given to them by instructors (Careless, 2006; Orsmond et al., 2005), many have difficulties in learning from it. Duncan (2007) identified that students often do not perceive that they could use feedback given on one assignment on later work, while Careless (2006) found that because instructor’s comments were too assignment specific, students could not use provided feedback in other work. Crisp (2007) ascertained that one-way feedback from instructor to student rather than dialogue could be the reason students fail to respond to feedback.

Crooks (1988) questioned whether feedback and summative evaluation were compatible and whether the feedback and summative purposes of student assessment are best separated. He felt that strong arguments for such separation had been presented by McPartland (1987), Miller (1976), Sadler (1983), and Slavin (1978), among others. They argued that when assessments counted as a high percentage of a student's final grade, the student tended to pay less attention to the feedback, and thus learned less from it. Crooks (1988) felt that this effect will be reduced if students are given multiple opportunities to test and prove their achievement, with only the final assessment counting toward their grade.

Black et al. (2003) argued that tests designed primarily to serve a summative function can draw out evidence of student achievement, and if used appropriately, can prompt feedback that increases student learning. This can also communicate to learners what is and is not valued in a particular discipline, thus communicating criteria for success. This approach opens up the possibility of students helping one another, and then using the tests as a guide to plan their own revision.

An increasing volume of literature advocated for the formative use of summative assessment. A focus on the purpose of an assessment, or how it is to be used rather than its timing was necessary to accept this view which blurs the distinctions between summative and formative assessment that many were comfortable with and wished to retain. The relationship

between graded assessments, summative or formative, and student motivation completes the review of literature related to the formative use of summative assessment from the instructor's perspective.

Using Grades to Change Student Motivation to Learn

The literature on motivation acknowledges the relationship between graded assessments and student motivation (Laurillard, 1997; Ramsden, 1992; Rust, 2002; Snyder, 1973). Walvoord and Anderson (2010) noted that “grading affects how students study, what they focus on, how much time they spend, and how involved they become in a course” (p. 2). Whereas grades are often used to control student behavior (Rowntree, 1987), a formative approach can counteract student's obsession with grades and redirect interest towards learning and, if students see summative assessment as formative, they will accept it more (Black et al., 2004). Learning-oriented students are more engaged and develop a deeper understanding of subject matter (Bloxham & Boyd, 2007). Bloxham and Boyd reported that:

Anecdotal experience tells us that, to a large extent, assessment activity in higher education is the learning activity. Students may take notes in lectures, seminars or from their reading, they may have been through the prescribed activities in laboratories or on field trips, but it is only when faced with assessment tasks that the majority seriously engages with that material. Tutors despair of trying to persuade students to undertake study which does not contribute in some way to their grades (p. 3).

Gibbs (1988) and Snyder (1973) found a significant amount of evidence that assessment affects students' learning focus. Students orient their learning towards what is being assessed. Both felt that assessment should be integrated into the learning process in such a way that the total learning experience was assessed (Trotter, 2006). Rowntree (1987) concurred with both Gibbs and Snyder that assessment could be used to encourage student learning.

Assessment can be used as an instrument of coercion, as a means of getting students to do something they might not otherwise be inclined to do ... Some teachers consider it as much a necessary part of their duties to supply students with motivation as it is to supply them with objectives and structured lessons ... Consider the teacher whose aim is for the student to become autonomous enough to develop his own goals and learning strategies.

Even he may feel that the student's motivation will be all the better for some external stimulus from assessment. Many such a teacher, while valuing his own freedom as to how he spends his time, will admit how the occasional deadline or external stimulus like the need to deliver a lecture, or prepare a report for a committee, can concentrate and energise his activities (Rowntree, 1987, p.379).

Studying the effect of continuous summative assessment on students in an undergraduate course in the United Kingdom, Trotter (2006), concluded that "the submission of regular work for assessment is welcomed by students and is effective at motivating them to study" (p. 512). This finding concurred with the findings of Elton and Laurillard (1979), who earlier found that even when an assessment was worth only a small amount of the total grade, students felt rewarded by receiving a grade for their work. Overall, students said that they spent more time on assignments that contributed towards their final grade and felt that the grade was a reward for their hard work. Regular submission of work for assessment motivated students to work throughout the semester, causing them to spend more time on their assignments than they would have otherwise (Trotter, 2006). All students indicated that they received useful feedback from the activities of preparing and submitting work for assessment.

In 2002 and 2004, The EPPI-Centre (Evidence for Policy and Practice: Information and Co-ordinating Centre), part of the Social Science Research Institute of Education, University of London, conducted a review of research on the effects of the use of summative and formative assessment for the same purpose (Harlen, 2005). Looking at the impact summative assessment has on student motivation for learning, Harlen and Crick (2005) found that when preparing students to pass high-stakes tests becomes the focus, teachers resort to a transmission style of teaching (Black & Wiliam, 1998; Linn, 2000; Stiggins, 1999), and that students can be trained to pass any type of test, even ones designed to assess higher levels of thinking. When teachers participated in the development of criteria used on summative assessments, they were more able to use them reliably when evaluating students' work (Hargreaves et al., 1996; Harlen, 2004a). Looking at assessments used for internal purposes, there was evidence that feedback from earlier assessments impacted the amount of effort students applied for future tasks of the same kind (Brookhart & DeVoge, 1999; Carter, 1997/8, Harlen, 2004b). There was also evidence that teachers changing assessment processes, practices and explanations they give to students can lead to better student learning (Flex et al., 1995; Harlen, 2004b).

Walvoord and Anderson (2010) made no distinction between formative and summative assessment arguing instead that “grading infuses everything that happens in the classroom” and that instead of trying to pretend that grading doesn’t affect students, educators should “use its power for student learning” (p. 1). They described grading as a complex context-dependent process by which a teacher uses classroom tests and assignments to shape student motivation with the aid of feedback and adjustments to teaching methods (Walvoord & Anderson, 1998; 2010).

Based on a study of 50 education lecturers (instructors) at an English university, Taras (2008) posited that due to problems of inconsistencies in the literature, lecturers do not understand the relationship between summative and formative assessments. Responses to survey questions indicated that lecturers are not sure about the relationship of formative assessment to grading and the role of summative assessment when it is not the ‘final’ summation). She concluded her study with several recommendations for future study including: “Future research needs to ask how we can reconcile formative and summative assessment so that they are mutually supportive: examining the processes of assessment seems one possible way of doing this” (Taras, 2008, p. 189).

Norm-Referenced and Criterion-Referenced Approaches

Bloxham and Boyd (2007) contended that differences in instructor approaches to grading assessments, categorized as norm-referenced or criterion-referenced, often result in unintended outcomes. Norm-referenced assessment is designed to distribute student performance over a range, for the purpose of discriminating between those who do well and those who do less well (Bowden & Martin, 1998). An example of this would be if 30% receive an A, 60% a B and the remaining 10% a C. Criterion-referenced assessment judges students against a set of criteria that is linked to desired learning outcomes. Bloxham and Boyd (2007) maintained that “whereas with criterion-referenced assessment all students have an opportunity to do equally well, a norm-referenced approach will almost always create a distribution of grades.” (p.82)

Price (2005) argued that because it is fairer to students, criterion-referenced assessment is generally considered the better of the two approaches. Students deserve to know assessment criteria beforehand and be judged based on the quality of their work rather than have their

performance compared to their classmates. Instructors benefit from a criterion-referenced approach because it gives them a way to justify their judgments (Sadler, 2005).

Sadler (2005) identified four different models of criterion grading used in higher education. Bloxham and Boyd (2007) adapted Sadler's model noting that models one and four represent the trend in universities:

1. The criteria are designed to judge how well the student has demonstrated progress towards the desired learning outcome.
2. 'Percentage grading' is used.
3. Same as two but allows staff to combine scores.
4. Involves specifying qualitative criteria or attributes. (pp. 82-83)

Educators feel that norm-referencing is still being used because it is easier and some grading schemes can only be understood and applied by referencing student work (Price, 2005). Bloxham and Boyd (2007) concluded that applying standards is not straight forward and requires contextual interpretation. Instructors in higher education use a combination of judgment and the application of grading standards criteria which are greatly influenced by the norms of the institution.

Standardized Tests

According to Guba and Lincoln (1989), tests have been used for hundreds of years. The earliest of these tests were designed to measure content mastery. The tests were usually given orally, one student at a time. If tests were of the written variety, the questions required essay type responses; this process was time-consuming and subjective. As the number of people being educated increased, such a system had to be modified to increase efficiency and objectivity.

In the United States, the term "standardized test" usually refers to a multiple-choice standardized achievement test. These centrally produced tests come with normative data that help give meaning to the test scores (Thorndike, 2005). The term "high-stakes test" refers to the use achievement of tests as the primary or only basis for decisions having major consequences. (Fraenkel & Wallen, 2009). Wiliam (2000) offered the following insight:

However, after a moment's reflection, it is clear that the contents of standardised tests and examinations are not a random sample from the domain of interests. In particular, timed written assessments can assess only limited forms of competence, and teachers are quite able to predict which aspects of competence will be assessed. Especially in 'high-

stakes' assessments, therefore, there is an incentive for teachers and students to concentrate on only those aspects of competence that are likely to be assessed. Put crudely, we start out with the intention of making the important measurable, and end up making the measurable important (p. 2).

Thorndike (2005) warned that these types of standardized tests are not likely to provide teachers useful information. He concluded that: (1) day-to-day instructional decisions should depend primarily on locally constructed rather than standardized tests, (2) grading decisions should be based primarily on locally constructed tests covering what has been taught in a given unit or course, but (3) diagnostic and remedial decisions can be based on information both from commercially produced diagnostic tests and from locally produced tests.

The U.S. Congress, Office of Technology Assessment (1992), defined a standardized test as one that uses uniform procedures for administration and scoring. Similarly, Thorndike (2005) and Walvoord (2010) note that the basic form of standardized refers to the use of uniform or standardized administration procedures. Therefore any test can be standardized if the conditions under which it is given are controlled and if identical scoring mechanisms are used for each group who takes the test. This includes oral examinations, essay writing, multiple choice tests, and performance-based tests (Butler & McMunn, 2006). In this study the term "standardized," when used, meant "uniform."

Standardized tests are usually classified as norm-referenced or criterion-referenced (Butler & McMunn, 2006). Norm-referenced scores, like those from the Nelson-Denny Reading Test, compare student performance to the performance of a normal group of peers, either national or local. Criterion-referenced or standards-based scores, like those from the use of a rubric, tell the teacher how well students are performing in terms of specific goals or standards. Norm-referenced tests cover a larger breadth of a specified domain and in less detail than a criterion-referenced test which focuses more directly on instruction (Fraenkel & Wallen, 2009; Thorndike, 2005).

Although the assessments used by DTAC instructors in this study could be classified as standardized assessments using definitions above, it was more exact to define them as common assessments. In addition to being standardized because they are uniform, the assessments used by

DTAC instructors were developed by the teaching organization itself instead of a higher or outside organization.

Teaching to the Test

“Teaching to the test” and “teaching the test” are terms used to describe ways instructors prepare students for summative assessments. These techniques are best described in terms Popham (2001) used, “curriculum teaching” and “item-teaching.” Teaching the test, or item teaching is considered unethical because it misrepresents how much students really have learned about a topic and results in unreliable data being used to make decisions at higher levels. Item-teaching is linked with “high-stakes” tests, like those used to evaluate the effectiveness of a school’s teachers.

The use of the test’s actual test items for practice or items so similar that they are almost indistinguishable from test questions is item-teaching, and this results in deceptive outcomes. If an instructor uses item-teaching and the skills for a particular block of instruction that CGSOC students are expected to master are not all represented on the assessment, then the results will not accurately represent student mastery. A teacher who chooses to item-teach is teaching a sample of skills chosen to represent a larger number of skills, and the outcome will be treated as representative of the student’s mastery. A perfect score would infer that a student can satisfactorily do problems of same the type but in a different context, which may not be true (Popham, 2001). This misrepresents how much a student really knows about a topic.

The use of broad learning objectives in curriculum design can lead to more item-teaching than to the learning objective, if assessments are not aligned with the curriculum. Resnick and Zurawsky (2005), contended that “When teachers match their teaching to what they expect to appear on state tests of this sort, students are likely to experience far more facts and routines than conceptual understanding and problem-solving in their curriculum....Narrow tests may not serve simply as the floor, but can become the de facto curriculum” (p. 11). Instructors who have their students prepare specifically for test questions are using time that could be used to teach the full curriculum.

The Grading Process

Educational practice is contextual in nature and activities need to be adapted to the classroom environment (Laurillard, 2002; Suskie, 2009; Walvoord & Anderson, 2010). This

means the responsibility falls on instructors and teaching teams to “critically evaluate assessment processes and procedures as they are used and developed within their local context” (Bloxham & Boyd, 2007, p. 7). Ultimately, the responsibility for evaluating these student officers falls on the individual instructor in the classroom who is charged with the dual, sometimes competing tasks of evaluating and developing student learning (Ewell, 2002).

Walvoord and Anderson (2010) described what they considered to be the grading process. Nationally known for leading hundreds of workshops on assessment and teaching-learning (including several at Ft. Leavenworth, KS), Walvoord contended that much of the need to write her book came from the recognition that the most requested workshop topics chosen by faculty participants were all grading related (2010). Walvoord and Anderson (2010) posited that grading is not just the act of grading assignments; it is a complex process that to be effective, integrates grading, learning, and motivation. They described grading as a complex context-dependent process with the goal of improving student learning:

By “*grading*” we mean not only bestowing an “A” or a “C” on a piece of student work. We also mean the process by which a teacher assesses student learning through classroom tests and assignments, the context in which good teachers establish that process, and the dialogue that surrounds grades and defines their meaning to various audiences. Grading encompasses tailoring the test or assignment to the learning goals of the course, establishing criteria and standards, helping students acquire skills and knowledge they need, assessing student learning over time, shaping motivation, planning course content and teaching methods, using in-class and out-of-class time, offering feedback so students can develop as thinkers and writers, communicating about student’s learning to appropriate audiences, and using results to plan improvements in the classroom, department, and institution (Walvoord & Anderson, 2010, p. 1).

Walvoord and Anderson (2010) felt that, “grading infuses everything that happens in the classroom. The grade is not an isolated artifact slapped on at the end; it must be integrated with everything else that happens in the classroom. It is a part of a process that includes shaping goals and assignments, communicating with students, helping them learn what they need, responding to them, and evaluating the quality of their work. Trying to establish an institutional assessment program unconnected to the grading process is wasteful” (Walvoord & Anderson, 2010, p.1).

Suskie (2009) agreed with Walvoord and Anderson's definition that the grading process is part of assessment; "Assessment, especially the grading process, motivates students to do their best" (p.59). Reeves (2011) noted that grading is a form of feedback, but it is also a very powerful instructional technique when it comes to influencing student achievement.

Goal Orientations

Chappuis (2015) discussed goal orientations as the key to understanding different student motivations answering the question, "Why am I doing this assignment?" (p.15). Student goal orientations fall into three categories (Ames, 1992; Schunk, 1996; Black & Wiliam, 1998a; Halvorson, 2012a):

1. *A learning orientation*, where the goal is to get better. When faced with difficulty, they get more involved. Motivated to learn and a willingness to engage in the process of learning. Have an intrinsic value of learning. They tend to seek help frequently while developing competence and then avoid help once mastery is perceived.
2. *A performance or ego orientation*, where the goal is to prove ability or hide a perceived lack of ability. Want the recognition of others and to be seen as smart. Achieving success with least effort is their methodology, leading to reluctance to engage in effort-based learning activities. When faced with difficulty, they exhibit anxiety and poor performance, sometimes leading to cheating. They tend to avoid seeking help to hide their perceived lack of ability.
3. *A task-completion orientation*, where the goal is to get it done and get a grade. Not interested in learning and mastery. Will expend just enough effort to get assessment turned in. When faced with difficulty, looks for ways to get completion, not understanding.(Chappuis, 2015)

Summary

This chapter provided a thorough review of the pertinent literature related to the formative use of summative assessments from the instructor's perspective organized into the following sections: (1) background of assessment, (2) formative use of summative assessment, (3) formative use of feedback, and (4) grading and student motivation. The literature on assessment theory supported the use of the same assessment for both summative and formative

purposes. Research at primary and secondary education levels showed that the use of graded assessments affected the activities instructors choose to use in an attempt to improve student learning. Literature on how the use of graded assessments in higher education affects activities instructors chose to use to develop student learning was not fully developed.

Chapter 3 - Methodology

The purpose of this case study was to explore if there were formative ways common summative assessments of student learning were being used and if they resulted in similar instructor practices in one teaching department of the US Army Command and General Staff College. This chapter is organized into the following sections: (1) research questions, (2) research design, (3) setting and participants, (4) data collection, (5) data analysis, (6) trustworthiness, (7) background and the role of the researcher, and (8) ethical considerations.

Research Questions

This study sought to answer the following research questions:

Research Question One.

Are there formative ways Department of Army Tactics instructors use common summative assessments?

Research Question Two.

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

The research questions were used to guide semi-structured interviews with participants. The full list of interview questions is located in Appendix E.

Research Design

A qualitative study approach was selected as the research method for this study because of the need to understand the context or setting in which instructors use assessments of student learning (Merriam, 1998). Qualitative studies are exploratory or descriptive in nature and stress the importance of setting, context, and participants' frames of reference (Marshall & Rossman, 2011). Patton (1985) explained that qualitative research "is an effort to understand situations in their uniqueness as part of a particular context and the interactions there" (p. 1). Merriam (1998) addressed the design of this study directly by explaining that the main concern

of qualitative research is to gain an understanding of the phenomenon from the participant's perspective, not the researcher's.

Creswell's (2007) comparison of five different approaches to qualitative inquiry strongly influenced the decision that a case study approach was the most appropriate methodology for this study because of the need to study the activities of one teaching department as they are engaged in the assessment process. Case study research involves the in-depth exploration of an activity or process (Yin, 2009), and individuals in a bounded system (Stake, 1995). The term "bounded system" is used because the case or issue being explored is bounded by time and place and has interrelated parts that form a whole, forming what Stake (1995) described as a "bounded system." Using this definition, this study was bounded by a one-month period of data collection with participants from the Department of Army Tactics teaching department, part of the Command and General Staff College, located at Fort Leavenworth, Kansas. Quantitative measures like the 2010 Command and General Staff School (CGSC) Faculty Survey can identify that instructors have a problem with assessments of student learning, but they cannot provide a deeper understanding of the effects of using assessments or help explain why instructors chose to use certain activities as a result of the assessments they are required to use (Creswell, 2007).

Setting

To help the reader understand the context of this study, this section describes the setting in which the data collection was conducted in June 2013. The study location was the U.S. Army Command and General Staff College (CGSC), located at Fort Leavenworth, Kansas. The CGSC is accredited by the Higher Learning Commission (HLC) of the North Central Association of Colleges and Schools as a Master's Degree Granting Institution. The college also sustains accreditation credentials to deliver Joint Professional Military Education (JPME) and Army Professional Military Education (PME). CGSC encourages free thought and the free flow of ideas and encourages aggressive examination of all academic subjects. As an institution accredited by the North Central Association of Colleges and Schools, CGSC subscribes to the American Association of University Professors 1940 Statement on Academic Freedom (*CGSC Catalog*, 2012). Although outside of the bounded time for this case study, in March of 2015 CGSC became a part of the newly created Army University.

The 10-month CGSOC is a graduate-level course that provides intermediate professional education for mid-career (majors and lieutenant colonels) from all military services of the United States as well as over 90 countries around the world. (*CGSC Self-Study Report*, 2005). In recent years students from 15 other US Government Agencies such as the Department of State, National Geospatial and Intelligence Agency, Federal Bureau of Investigation, Border Patrol, and the Department of Homeland Security with limited military background have also been invited to attend.

Within the Command and General Staff College (CGSC), the faculty members of the Command and General Staff School (CGSS) are responsible for the education of over 1300 student officers each year at the resident (Fort Leavenworth campus) 10-month Command and General Staff College Officer Course (CGSOC). In Academic Year 2013 (AY 2013) this included 115 international students (Davis & Martin, 2012). The Department of Army Tactics (DTAC) is one of six departments that compose CGSS. DTAC provides a program of instruction throughout the Army that grows leaders to plan, execute, and command in Decisive Action and combined arms operations for a campaign-quality Army with joint and expeditionary capabilities (CGSC 350-1, Sep 2012).

Participants in this study were a mix of civilian and active duty military instructors from the Department of Army Tactics (DTAC), a part of the U.S. Army Command and General Staff School (CGSS). Civilian instructors were mostly retired U.S. Army colonels and lieutenant colonels that were hired to teach as Department of the Army Civilians. Active duty instructors were lieutenant colonels and senior majors mostly from the Maneuver, Fires and Effects (MFE) branches such as Armor, Artillery, Aviation, Infantry, Military Intelligence, and Aviation.

CGSC classes described in this study were taught at the Fort Leavenworth Lewis and Clark Building, completed in 2007. Each state-of-the-art classroom sat up to 16 students (the normal class size) and had two large, flat-screens; high-definition television sets that served as multi-functional displays (Figure 3.1). Dry erase white boards were mounted on three of the walls and networked computers were built into each student desk (CGSC 350-1, Sep 2012).

Figure 3.1 CGSC Classroom at the Lewis and Clark Center



Figure 3.1. Command and General Staff School students in a Department of Army Tactics Department class, in the Lewis and Clark Building, Fort Leavenworth, KS. Used with permission: U.S. Army Photo by Dan Neal.

Bloxham and Boyd (2007) wrote that instructors that teach standardized curriculums that use standardized assessment instruments lack control over scheduling and are limited in the teaching practices they can employ. This was true of the DTAC instructors in this study who taught a standardized (uniform) curriculum and used common graded assessments developed by DTAC curriculum developers and instructors. At the explicit direction of the CGSC Deputy Commandant, faculty members enjoyed broad latitude to adopt distinctive approaches in the classroom or reconstruct lessons on their own (*CGSC Self-Study Report, 2005, p. 101*). Graded

assessment instruments, including due dates, could not be changed without the consent of the department director (DTAC Standing Procedure) because data collected from these direct measures was part of the CGSC program evaluation process and was used to make program improvement decisions (CGSC Bulletin #8, 2010). In the AY 13-01 Intermediate Level Education (ILE) Core Course and Advanced Operations Course, DTAC instructors were responsible for four graded assessments for each student; one written analysis and three for contribution to group exercises (C500, O300, and O400 Assessment Plans, AY 13-01).

Population

The population for this study was Department of Army Tactics instructors assigned to the Command and General Staff School. During the data collection period (June 2013), there were 122 instructors assigned to the Department of Army Tactics who taught a centrally developed common core curriculum. This number includes both active duty military and Department of the Army civilian instructors who taught either an August or February start class at the Fort Leavenworth campus and those who taught at one of four satellite locations: Fort Belvoir, VA; Fort Gordon, GA; Fort Lee, VA; and Redstone Arsenal, AL. (See Table 3.1 for characteristics of the total population).

Table 3.1 Population Characteristics: All DTAC Instructors (AY 2013)

Table 3.1 Population Characteristics: All DTAC Instructors (AY 2013)		
Characteristic	Quantity	Researcher Comment
Civilians at Ft Leavenworth	77	Department of Army Civilians (DAC)
Active Duty Military at Fort Leavenworth	25	Major/ Lieutenant Colonel
Civilians at Satellites	13	Department of Army Civilian
Active Duty Military at Satellites	7	Major/ Lieutenant Colonel
Note: Data for the June 2013 period of data collection. Population of all DTAC instructors was 122.		

Table 3.1. All assigned DTAC instructors for the case study time window.

With the intent of achieving maximum variation, participants were purposefully selected from among the total population of 102 Department of Army Tactics instructors that taught at the Fort Leavenworth campus during the data collection period (June 2013). Although instructors at the satellite locations taught the same curriculum, the use of Video Tele Conferencing would not have provided the level of intimacy desired when trying to obtain honest and candid answers to personal questions and the cost of travel to conduct face-to-face interviews was prohibitive.

Sampling Strategy

In qualitative research, sample sizes are typically small and the participants purposefully selected for their ability to provide detailed information on the topic studied. This study used purposive or purposeful sampling to select participants. Patton (1990) noted that “qualitative inquiry typically focuses in depth on relatively small samples...selected purposefully” (p. 169). Merriam (1998) agreed offering that “purposeful sampling is based on the assumption that the investigator wants to discover, understand, and gain insight and therefore must select a sample from which the most can be learned” (p. 61). Fraenkel and Wallen (2009) described purposive

sampling as “a nonrandom sample selected because prior knowledge suggests it is representative, or because those selected have the needed information” (G-7).

Study participants, including focus group members and pilot interviewees came from the total population of 102 instructors assigned to the Department of Army Tactics (DTAC) who taught at the Fort Leavenworth campus (see table 3.2 below). Of these instructors, about one-half met the screening criteria; had taught the CGSOC Intermediate Level Education curriculum at least once, had been a course or lesson author, and had been a Staff Group Advisor. Representation from different teaching teams, active duty instructors, and civilian instructors was sought. Women, African-Americans, and International Exchange Officer instructors were under-represented in DTAC and were under-represented in this study.

The rationale for each criterion chosen is explained below:

1. Years teaching in DTAC: Participants needed to have taught the entire 10-month CGSOC curriculum at least once to gain an appreciation for the sequencing and interrelationship of the assessments of learning throughout the course -the more years teaching in DTAC the better.
2. Course or lesson author experience: These instructors have developed curriculum and common assessments of student learning and have an understanding of institutional policies governing each.
3. Civilian or Military: The majority of instructors (approx. 60%) were civilian. Representation equal to the percent of military instructors in DTAC was sought to include their insights.
4. Staff Group Advisor: Primary and Assistant SGAs have coached students about their academic progress and understand how the grading process works.
5. Teaching Team: Each teaching team had four DTAC instructors and its own unique culture and methods of faculty development. Representation from as many different teams as possible was sought.

Purposeful Sample

Ten instructors were selected for participation using the sampling strategy described above. This number was appropriate to achieve saturation, or the point when no new information

could be found that added to the understanding (Creswell, 2007). If saturation would not have been reached with the initial selections, research would have continued with the selection of more participants. Volunteers were solicited from the population pool individually or via personal e-mail and provided a written statement about the nature and purpose of the study in order to make an informed decision. The intent was to choose individuals that were “information-rich” (Patton, 1990, p. 169) and could aid in the understanding of the central phenomenon (Creswell, 2008). Participant characteristics are described in Table 3.2 below.

Table 3.2 Participant Characteristics: DTAC Instructors at Ft Leavenworth (AY2012-2013)

Table 3.2 Participant Characteristics: DTAC Instructors at Ft Leavenworth (AY 2012-2013)						
“Name”	Years Teaching in DTAC	Course Author Experience	Lesson Author Experience	Civilian or Military	Years SGA/ASGA	“Teaching Team”
“Aaron”	3+	no	yes	Military	2+	“D”
“Allen”	1 +	no	yes	Military	1	“J”
“Alex”	2 +	no	yes	Military	2	“B”
“Arlen”	3 +	no	yes	Military	3	“J”
“Caleb”	5+	yes	yes	Civilian	5	“G”
“Dexter”	5+	no	yes	Civilian	5	“A”
“Frank”	5+	yes	yes	Civilian	4	“H”
“Fred”	3+	no	yes	Civilian	2	“B”
“Hank”	4+	no	yes	Civilian	4	“L”
“Jack”	5+	no	yes	Civilian	5	“C”
<p>Notes:</p> <ol style="list-style-type: none"> 1. Pseudonyms for military instructors start with “A” for easy recognition. 2. Numbered Teaching Teams were randomly assigned letters. Nine of ten instructors were on different teams. 						

Table 3.2. Does not include the two pilot participants.

Procedures

Data Collection

Marshall and Rossman (2011) explained that “Qualitative researchers typically rely on four primary methods for gathering information: (1) participating in the setting, (2) observing directly, (3) interviewing in depth, and (4) analyzing documents and material culture, with varying emphasis” (p. 137). Data collection for this study emphasized in-depth interviewing. Curriculum documents to include common assessments of student learning that were created by instructors were used to help understand why instructors chose to use specific activities to improve student learning.

After the study participants were identified, data collection activities were sequenced to facilitate refinement of the interview questions. First, a focus group interview was conducted to help focus the study, and then pilot interviews were conducted to reduce errors, followed by research interviews for collection of the primary research data. Members of the panel of experts were asked to review the interview questions as they are being developed and revised to ensure continued focus on the purpose of the study and provide advice on sentence structure and clarity.

Figure 3.2 Sequential Flow of this Qualitative Case Study

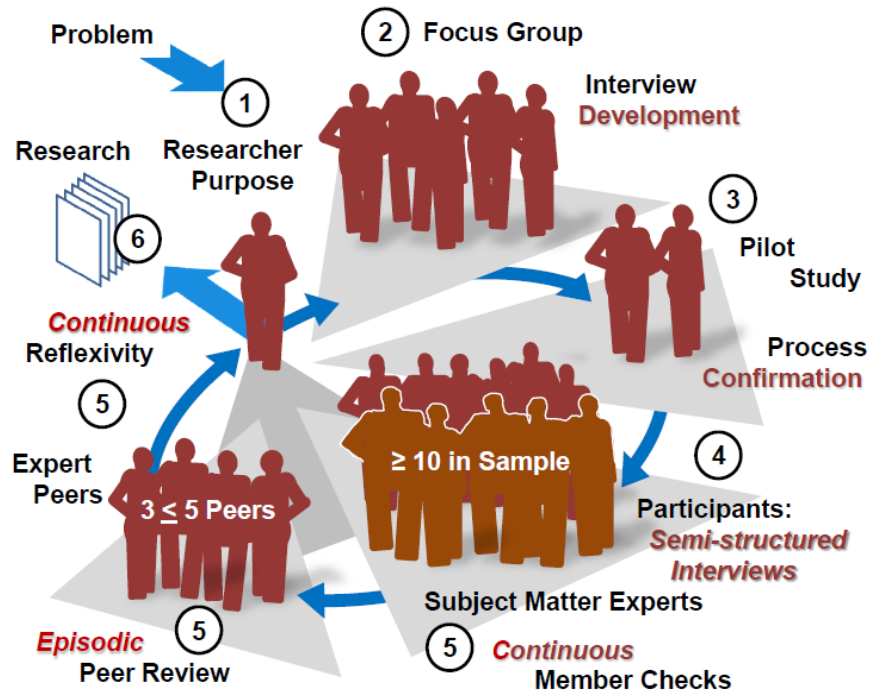


Figure 3.2. This illustration is a sequential flow through the six main phases of qualitative exploration using bounded case study methodology as in this research study. As adapted from Creswell (2008), *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, Upper Saddle River, NJ: Pearson Education (3rd ed.), p. 52. Original art used by permission of Jon H. Moilanen.

Focus Group

The researcher facilitated a one-hour focus group interview with four Department of Army Tactics instructors to understand the range of views, help focus the study, and aid in refinement of the interview questions found in Appendix A, Interview Protocol (Bogdan & Bilken, 2007). Based on the experiences of Guba and Lincoln (1981), three to four participants should be selected because a group of this size minimizes the possibility of a situation where several people try to talk at once and makes it easier for the facilitator to prevent more vocal members from dominating the discussion. The focus group participants were volunteers purposefully selected by the researcher because they were representative of study population

demographics. Instructors who had been teaching at the Command and General Staff College for over five years and had a graduate degree in adult or higher education were sought. These four individuals did not participate in either the pilot interviews or research study.

The focus group purpose and sequence of events were clearly explained to group participants by the researcher to focus discussion and achieve session objectives. A short introduction included a brief overview of the research study, the objectives of the focus group session, guidelines, and confidentiality considerations for focus group discussion participants. The semi-structured interview questions developed for study participants were used to frame discussion and keep the group focused. Prior to conducting the focus group interview, participants read and signed the Informed Consent Form (see Appendix B).

Pilot Interviews

Two pilot interviews were conducted with the goal of improving the interview process and further refining the individual questions derived from the focus group session. Questions were checked for clarity and usefulness in eliciting relevant data (Creswell, 2007; Yin, 2009). The pilot interview participants were one active duty officer and one Department of the Army civilian purposely chosen from the population.

The pilot interviews were conducted as a full dress rehearsal and included all steps of the process; recording, transcription, panel of expert review of the results and member checking. These expert peers were members of the Command and General Staff College faculty, who have completed or are working towards completion of a doctorate. Additionally, the pilot provided an opportunity to use the NVivo10 qualitative computer software package to help code, organize, and sort data. Prior to conducting the pilot interviews, participants read and signed the Informed Consent Form (see Appendix B) and the Instructor Survey Form (see Appendix C).

Semi-structured Interviews

Face-to-face semi-structured individual interviews scheduled for one-hour were the primary means of data collection for this study. Creswell (2007, 2009) and Merriam (1998), provided the basis for development of the interview protocol (Appendix A) which was reviewed by expert panel members. The open ended nature of the interview questions facilitated the ability to build upon participant responses to complex issues in order to explore them more thoroughly. Prior to conducting the interviews, participants read and signed the Informed Consent Form (see

Appendix B). The interviews were conducted in private locations, free from distractions (Creswell, 2007). The researcher's familiarity with the study location facilitated confidential meetings.

Before the interview questions were asked, participants were asked to look at the Common Assessment Classification Worksheet (see Appendix D) and identify from a list, which of the DTAC assessments taught in AY13-01 they would classify as summative. This purpose of this exercise was to ensure that the participants were using the same definition for summative assessment as the interviewer. The mutually understood definition was, "summative assessments measure what students have learned at a chosen point in the curriculum against a standard." Copies of each of the assessments were available for reference but not needed because of participants familiarity with the assessments.

Interviews were tape recorded and professionally transcribing was used to convert digital recordings of the interviews into Microsoft Word documents. The transcriber signed a statement of non-disclosure. The interview transcripts were member checked for accuracy and precision. During the data analysis phase of the research the transcripts were peer reviewed. Interview and field notes will be maintained for five years following the completion of this study. Pseudonyms were used in all written material to identify participants and to provide participant anonymity. Descriptions of participants and identifying references were written in a general manner when needed to respect confidentiality.

In addition to interviews, documents used included the Researcher's reflective journal, AY 2013 Command and General Staff School Intermediate Level Education Core Course curriculum and lesson plan materials. These lesson plans provided the assessment plan for each course to include common assessment of student learning instruments created by DTAC instructors.

Data Analysis

Data analysis was based on procedures designed to produce credibility, dependability, confirmability, and transferability (Guba & Lincoln, 1985). Hand coding as well as the qualitative computer software package NVivo10 was used to code, organize, and sort data. Input from independent coders combined with member checking of the transcripts by participants increased objectivity and helped mitigate researcher bias. Episodic peer debriefing (also called

peer review) involving several expert peers was used during data analysis to enhance the accuracy of the analysis (Creswell, 2009). Data analysis procedures generally followed the six-step process described by Creswell (2009, pp. 185-190). All data was organized by source type and then read through to obtain a general sense of the overall meaning. Detailed analysis began with coding based on key words, the literature, and common sense. The coded data was then analyzed to generate themes and descriptions that were further analyzed to determine how best to represent the results. The final step was making an interpretation or meaning of the data. Meaning was derived by comparing the findings with information derived from the literature reviewed and looking for conformation or divergence.

Trustworthiness

Quantitative research is judged on the criteria of reliability, validity, objectivity, and generalizability (Creswell, 2009). Guba and Lincoln (1985) broadened these criteria for use with qualitative research by advancing the alternative constructs of credibility, dependability, confirmability, and transferability known collectively as “trustworthiness.” Marshall and Rossman asserted that research grounded in trustworthiness ensures the rigor and usefulness of a qualitative study (2011).

Creswell (2009), based on Guba and Lincoln’s (1985) writings, equated trustworthiness with qualitative validity, a determination that findings are accurate from the standpoint of the researcher, reader and participant. This study achieved trustworthiness by implementing multiple validity strategies recommended by Creswell (2009): (1) triangulation of data sources, (2) member checking, (3) use of rich, thick description to convey findings, (4) clarification of researcher bias, (5) inclusion of negative findings, (6) prolonged time in the field by the researcher, (7) peer debriefing, and (8) use of an external auditor (p. 191-192). Use of these strategies to check the accuracy of the findings is discussed in the following paragraphs.

Triangulation was achieved by the use of different data sources of information, adding to the validity of the study. The convergence of reviewed literature with perspectives from participants and experts was used to identify the themes and sub-themes. Member checking was used to ensure accuracy throughout the research by having participants review transcripts and themes to see if they felt they were accurate.

Rich, thick description was used to convey the findings, presenting an element of shared experience, adding to the validity of the study. The findings were presented in a rich and descriptive manner, showing the different participant perspectives and providing realistic insight into a unique educational culture. Researcher bias was a concern because the researcher was a peer of the participants. The role and background of the researcher was explained in detail to clarify the bias that the researcher brought to the study.

Evidence and perspectives that ran counter to the themes and sub-themes were presented to make the account more realistic and valid. The researcher had spent prolonged time in the field and had an in-depth understanding of the research topic. The researcher's experience with the participants in the actual research setting contributed to more accurate and valid findings. Peer debriefing was used to enhance the accuracy of the findings and add validity. This involved peer and expert debriefers who reviewed themes and findings and asked questions about the study so that the account would be understood by people other than the researcher. The use of an external auditor that was not familiar with the researcher or the study as described by Creswell (2009) was not a part of the research. Supervisory Committee Members performed the auditor role by reviewing the project and providing an objective assessment during the process or at the conclusion of the study.

MacNaughton (2001) advocated that in qualitative research, the aim is to make the process of data analysis as visible as possible. Ortlipp (2008) explained that in qualitative research, the goal is to make visible to the reader the constructed nature of research outcomes, what Mruck and Breuer (2003) described as a construction that "originates in the various choices and decisions researchers undertake during the process of researching" (p. 3). A reflexive approach in the form of a reflective research journal was used by the researcher to add transparency to the research process. Starting with data analysis, episodic entries were added to the research journal enabling the researcher to revisit and draw on past thoughts and construct outcomes that were presented in the final chapter as discussion, implications, recommendations, and reflections.

Role and Background of the Researcher

In qualitative research, the researcher is the key instrument for collecting and analyzing the data (Creswell, 2009; Merriam, 1998). Guba and Lincoln (1981) noted that certain

characteristics differentiate the human researcher from other data collection instruments: the researcher is responsive to the environment; he or she can adapt techniques to changing circumstances; the total context can be considered and interacted with; what is known about the situation can be expanded through awareness of nonverbal aspects; the researcher can process data immediately, can study and summarize as the study evolves, and can explore anomalous responses. Thus the role and background of the researcher needed to be clarified to ensure the credibility of this study.

In applied fields, such as education, personal experience with and the interest in a topic often drive the researcher to conduct a study at their own workplace (Marshall & Rossman, 1999). Creswell (2007) noted that such a study may raise questions about whether good data can be collected when the act of collecting data may introduce a power imbalance between the researcher and participants. The requirement for a qualitative researcher, who is researching a familiar topic as is the case with this study, is to demonstrate that personal interest will not bias the study. The primary researcher for this case study gained approval from the CGSC for the study and permission to interview participants (Marshall & Rossman, 2006), because “he was one of them.” This is what Glesne and Pushkin (1992) described as doing research in your own “backyard” (p. 21).

Patton wrote that “The purpose of interviewing is to find out what is in and on someone’s mind” (p. 278). In this research setting the fact that the researcher was a peer to and has the foundation of rapport with those he interviewed was beneficial. Merriam (1998) explained that a researcher is better able to conduct an interview in an atmosphere of trust.

For this study the primary researcher was an educational leadership doctoral student employed as an instructor in the Department of Army Tactics at the U.S. Army Command and General Staff College. He had worked in this department for over eight years, as both an active duty military and civilian instructor and taught the same common curriculum as the selected participants. During his time as an instructor, he served as a lead instructor, course author, and lesson author. Having completed the Lesson Author’s Course, he developed and assisted in the development of common lessons and assessment instruments that were used during the 10-month Command and General Staff Officer Course (CGSOC). A Staff Group Advisor (SGA) for seven academic years, he coached and advised the students he taught and was responsible for writing

their annual evaluation reports. The researcher had no evaluative role over any of the participants.

Ethical Considerations

Guba and Lincoln (1981) voiced concern about what they called “unusual problems of ethics” (p. 378) associated with case study research because of the ease with which a researcher could select for use only data needed to illustrate a desired outcome. Glesne and Peshkin (1992), noted that readers needed to be aware that research conducted in a researcher’s own workplace or “backyard” research, has the possibility of personal bias affecting findings. These concerns were at the heart of the trustworthiness of this study and were mitigated by the qualitative validity strategies described previously.

Involvement in this study was voluntary. Potential participants were contacted individually or via personal e-mail and provided a written statement about the nature and purpose of the study in order to make an informed decision. Participants signed statements acknowledging their rights. Individuals who transcribed or reviewed data signed non-disclosure statements (Appendix D). All data, records, and field notes will be safeguarded in a secure location for at least five years to prevent public disclosure of interview responses. A copy of the informed consent form used in this study is enclosed as Appendix B.

The Kansas State University (KSU) Institutional Review Board (IRB) approved application and extension to conduct research with assigned tracking number is enclosed as Appendix A. The approved application to conduct research at the U.S. Army Command and General Staff College (CGSC), Fort Leavenworth, Kansas is enclosed as Appendix B. An institutional agreement was signed between KSU and the CGSC because of unique Department of Defense requirements for non-exempt research involving human subjects. CGSC agreed with the KSU IRB and acknowledged that the KSU IRB was the reviewing IRB.

Summary

This study used a qualitative case study research design that included semi-structured interviews as the primary source of data. The study was bounded by a one-month period of data collection with participants from the Department of Army Tactics (DTAC), part of the Command and General Staff College, located at Fort Leavenworth, Kansas. Purposive sampling was used to identify appropriate study participants. Data collection activities consisted of a focus group

interview, two pilot interviews and then ten research interviews. Data analysis was based on procedures designed to ensure trustworthiness (Guba & Lincoln, 1985). Interview transcripts were member checked for accuracy and precision. Episodic peer debriefing involving several expert peers was used to enhance the accuracy of the analysis and mitigate possible bias (Creswell, 2009).

Chapter 4 - Findings

Introduction

This chapter presents the findings of qualitative research concerning the formative use of common summative assessment by instructors in the Department of Army Tactics (DTAC) during the 2013 Academic Year (AY). Brief profiles of the instructors who participated in this research are provided. The findings from the semi-structured interviews conducted during data collection are presented without conclusions. Participant perspectives and experiences with common summative assessments are presented from their particular point of view. Two research questions guided this exploratory research on the formative use of common summative assessments:

Research Question One

Are there formative ways Department of Army Tactics instructors use common summative assessments?

Research Question Two

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

Qualitative Methodology

This study followed a qualitative case study methodology designed to facilitate in-depth exploration of a specific activity and individuals bounded by a determined period of time: the formative use of DTAC common summative assessments by selected AY 2013 Department of Army Tactics instructors who taught at the Fort Leavenworth CGSOC location, a one-month period of data collection during the month of June, 2013 and member review of transcripts.

Demographics

Ten DTAC instructors participated in this research (an additional two participated in the pilot interviews); all were selected by using the purposive sampling strategy described in Chapter Three. Participants were purposefully selected with the intent of achieving maximum variation

and exposure to different individual and team practices. Closely matching the demographics of the population of DTAC instructors at the time of data collection, sixty-percent of the participants were Department of the Army Civilians (DAC), and forty-percent were active-duty field grade officers. Ten of the sixteen AY 13-01 Command and General Staff Officer Course teaching teams were represented.

Participant Profiles

Profiles of the instructor participants are provided to show the level of experience each had with DTAC common summative assessments. All had earned a Master's Degree, had taught the AY13-01 DTAC curriculum and attended DTAC curriculum train-ups conducted prior to the conduct of each block of lessons. All had served as a lesson author which is important to note because lesson authors assisted course authors in the development of common summative assessments. All had served as a Staff Group Advisor at least once meaning that they would have coached student officers throughout the academic year on assessments and grades. While there were many similarities in the participant's backgrounds and experiences, the findings presented in this chapter show a wide range of individual perspectives.

Because of confidentiality concerns with this research, participant profiles do not include information that might compromise their anonymity, such as exact periods of service and Army branch (i.e. Armor or Infantry). As part of the interview protocol, all participants were assured of anonymity and each was assigned a pseudonym, shown in bold. The instructor profiles are presented in the order in which the individuals were interviewed, to include the two pilot interviews, which are the first two below.

“LTC Alfred” participated in the Pilot Interview. He was an active-duty Army officer who had served in DTAC for more than three years. He was a Team Leader and Lead Instructor, had served as a lesson author for more than three years and served as a primary or assistant Staff Group Advisor three times. LTC Alfred felt that summative assessments measure what a student has learned and they should test a student's ability to tie concepts together whereas a formative assessment should provide him instant feedback on whether a student, “got it.”

“Mr. Calvin” participated in the Pilot Interview. He was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than four years. He was a Lead Instructor, had been a lesson author for over two years, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor four times. Mr. Calvin described summative assessments as a summation of what students have learned at a certain point in time, which should summarize a block of instruction. For him the distinction between summative and formative assessment was that the latter provides feedback to students.

“Mr. Frank” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than five years. He was a Lead Instructor, had been both a course author and lesson author, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor more than four times. Mr. Frank didn’t see much difference between how he used summative and formative assessments. Describing a common summative assessment, he commented that, “officially it’s a summative, but I also use it for counselling, and I also look at it to see if there is anything I need to do to adjust my teaching.”

“Mr. Fred” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than three years. He had served as a lesson author for over one year and had served as a primary or assistant Staff Group Advisor more than two times. Mr. Fred, referring to the different populations of students attending CGSC that enter at different levels of military skills, emphasized that students should always be assessed on a skill first as a formative assessment, with feedback provided, before being assessed summatively.

“LTC Aaron” was an active-duty Army officer who had served in DTAC for more than three years. He had served as a lesson author for more than one year and served as a primary or assistant Staff Group Advisor more than two times. LTC Aaron expressed that for him, formative assessments are opportunities for students to dialog about their experiences and learn whereas summative assessments come at the end of a curriculum block, where instructors are essentially

asking students, “What have you learned? What do you understand? How well do you understand the concepts?”

“**Mr. Caleb**” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than five years. He was a Lead Instructor, had been both a course author and lesson author, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor more than five times. Mr. Caleb explained that, “although we understand the primary difference between formative and summative, ultimately it’s the feedback that we’re trying to provide, and also assess what [students] have learned.”

“**LTC Allen**” was an active-duty Army officer who had served in DTAC for more than one year. He had served as a lesson author for less than one year and served as a primary or assistant Staff Group Advisor once. LTC Allen looks summative and formative assessments in much the same way by using the results to determine what students have learned and, “was I way off the mark on that one, do I need to either comeback and re-tweak how I teach it next year, or figure out how to build it back into the rest of the work I’m doing [this year].”

“**LTC Alex**” was an active-duty Army officer who had served in DTAC for more than two years. He was a Team Leader and Lead Instructor, had served as a lesson author for more than one year and served as a primary or assistant Staff Group Advisor two times. LTC Alex explained how he saw the difference between formative and summative assessments by using the C500 Individual Exam summative assessment as an example, “... it’s at the end of a block [of instruction], so it doesn’t really give me the interactive, real-time assessment of where the students are as well as where I am as an instructor.”

“**Mr. Dexter**” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than five years. He was a Lead Instructor, had been a lesson author for over five years, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor more than five times. Mr. Dexter believed that summative assessment was when students demonstrated that they met a standard of knowledge,

understanding, or ability to communicate whereas formative assessment was, “free feedback,” that students should receive as an opportunity to improve themselves.

“**Mr. Jack**” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than five years. He was a Team Leader and Lead Instructor, had been a lesson author for over five years, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor more than five times. Mr. Jack expressed that he feels students want to be assessed, “they want to be told, no matter what they do, what my assessment of the quality of their work is.”

“**LTC Arlen**” was an active-duty Army officer who had served in DTAC for more than two years. He had served as a lesson author for more than one year and served as a primary or assistant Staff Group Advisor two times. LTC Arlen looked at DTAC summative assessments as being both summative and formative because students get feedback on how well they did and instructors get feedback that can be used to modify teaching.

“**Mr. Hank**” was a retired Army officer who had served as a Department of the Army Civilian in DTAC for more than four years. He had been a lesson author for over three years, was a graduate of the CGSC Course Author’s Course (FDP3) and had served as a primary or assistant Staff Group Advisor four times. Mr. Hank believed that summative assessment was something that is given after a formative assessment on the same skill, “to test them individually, at that point [in the curriculum].”

Qualitative Interview Results

Four major themes were identified during data analysis. Themes one and two are aligned with research question one, themes three and four are aligned with research question two. The interview results are presented by theme, and then categorized by the findings/subthemes within each theme.

- Theme One: Feedback Given to Students after a Summative Assessment
- Theme Two: Modification of Teaching and Learning Activities
- Theme Three: Preparation of Students for Upcoming Summative Assessments
- Theme Four: Use of the Grading Process by Instructors

Research Question One

In what formative ways do Department of Army Tactics instructors use common summative assessments?

Theme One: Feedback Given to Students

“Feedback is probably the most powerful communication that instructors can regularly use to affect learners’ competence” (Wlodkowski, 2008)

Instructor comments about feedback given to students after a summative assessment included the findings/subthemes below. The numbers in parenthesis indicate how many different instructors out of 12 participants (10 primaries plus two pilots) addressed the topic/subtheme:

- Written feedback was provided to students (10)
- Common errors/trends were discussed or provided to the class (10)
- Feedback was applicable to other assignments (8)
- Exemplars were used to show students “a way” (6)
- Individual coaching was provided to students (5)
- Timeliness of feedback affected usefulness to students (5)
- References were provided to students for further learning (4)

Descriptive feedback, with guidance on how students can improve, while learning is ongoing, is a key characteristic of formative assessment and one of the instructor practices that gives formative assessment its power. Although descriptive feedback is not a characteristic associated with summative assessment, participants in this study chose to provide students feedback in many different ways. The participants did not consciously chose to give students

feedback on their summative assessments, they chose to give feedback on every product students submitted, no matter the classification. Chappuis (2105) specified that for feedback to be effective, it is regular, descriptive and answers several questions for students, “What are my strengths?”; “What do I need to work on?” and “Where did I go wrong and what can I do about it?” Participant transcripts reported they provided feedback to students that answered these questions with the intent of improving student learning; formative use of common summative assessments. This section starts with findings on written feedback provided to students and then discusses in detail each of the other six ways common summative assessments were used formatively.

Written feedback was provided to students

Ten instructors provided written feedback on summative assessments, much beyond what was required on the CGSC 1002 Assessment Form, the identification of just right and wrong answers, or just summarizing learning at a specific point in time.

Frank explained that he always provided individual feedback in the form of written comments on summative assessments, “as I go through and I grade them, I write comments on the exams themselves. And I try to write them clear enough, but also be concise enough so that they [students] understand what the general points are.”

Frank continued that he felt there was no real difference between summative and formative assessments during the initial Classification Exercise. He explained how he had used an assessment classified as summative for providing feedback to students, “I mean officially it’s a summative assessment, but then I also use it to come back for counseling.” When asked if he thought that a particular assessment had been designed to be summative, he replied, “Yeah, I think the school designed it as summative, but I think you can... get a formative benefit out of it as well.”

Archie routinely provided feedback in the form of written comments to address student performance on their assessments. While providing feedback, if he thought a student had “missed the mark” on a particular question or a series of connected questions, at the bottom of their exam he wrote, “Please schedule time to see me. There was a limit to the feedback that could be provided on paper, higher level concepts required meeting with students in person to ensure they understood.”

Caleb quickly answered, "...when they get the summative assessment, I do it two ways. I'll give them specific feedback, I either hand write it, or whatever the [question] is, or I'll say, I think you're missing the point here, it's not in accordance with doctrine, or it may be another question that says, did you consider more integration or this aspect of, of transitions or integrations, and I do that individually."

Caleb felt strongly about providing his students feedback no matter the assessment classification, "although we understand the primary difference between the formative and summative [assessment], ultimately it's the feedback. And so it's the feedback, that we're trying to provide, and also assess what they've learned."

Allen was very detailed in the description of his actions, "Typically what we'll do is, first of all, I try to make lots of notes on the actual homework or the assessment that they've turned in. So give that back to them and ideally those have enough detail in them that they'll stand alone. Or at least give them a pretty good idea of what it was that I wanted them to get out of, or what's supposed to answer on the assessment. Then we'll typically walk through it in class, especially if there were questions that were giving a lot of them problems. Then as part of the 1002s that they get back, I typically have some extra notes that are put in there. Especially if it's somebody who is really struggling on something, I typically have some extra notes that are put in there."

Alex created his own evaluation sheet that was attached to each assessment providing individual feedback. "Depending upon how they graded out. If most of the students did well, and there did not seem to be any general trends, negative general trends, I will pass back and I give them their individual feedback with the exam. I generally do not use the 1009 that's associated with that. I will generally create my own, type it up and attach it, and give them individual feedback. "

Fred also provided his students with written feedback and when needed, had them come see him to for individual feedback, "I mean generally those comments will be, you know, things like, "Good Point," or "That's a new way of looking at it," or "You're not expressing your idea clearly." "If there's something that's obviously factually wrong, I'll mark that on there."

Jack felt that he provided lengthy individual written feedback. During the process of grading an assessment he "usually" gave substantive feedback that included a recommendation for how students could improve on the next assessment. Referencing the friendly course of

action statement and sketch summative assessment, Jack said that he concluded his comments with a statement that identified and addressed obvious shortfalls such as, “the next time you do this, concentrate in the following areas.”

Jack was positive that his students wanted to be assessed. “They want to be told, no matter what they do, what my assessment of the quality of their work is... for instance a group of three or four coming up with a course of action statement on the white board, they want a definitive assessment as to the quality of what they’ve produced or it’s deficiencies, and they want that from their instructor and their peers.” An experienced instructor with more than five years teaching in DTAC, Jack felt that not only did his students want to hear his evaluation of their work, they wanted to hear from their fellow students as well. He added “and not just a discussion”, they want it in writing.

Alfred (Pilot Interview) felt strongly about the importance of providing individual feedback. “... probably how I provide the most feedback is individually, I’ll even scratch things out and say, “This would have been a better way to say this”, or “Did you think about this?” or “Did you think about that?” Alfred further explained that unless sought out by a student about the way an answer was graded or a particular concept that they didn’t understand, he provided his feedback on their papers.

Calvin (Pilot Interview) was aware that his students may not have time in class to understand his comments, “I write on the document so they can read through it, they can see my comments. I give them time to digest that, and to understand what I was saying on their papers, whatever that assessment is. And then, because we don’t have the time ...I don’t take the time in class to say, “okay, here’s question one”, because there are so many different responses to it. I’d like them to have them come back to me if they have questions about something, to clarify what I wrote. Or if they disagree with what I wrote, then I give them the opportunity to come to me and address it with me.”

Dexter provided students written feedback by exception only. He explained, “Where there’s clearly a misunderstanding of something related to the learning objectives, I’ll make comments. Some [comments] are more extensive than others. If from my personal judgment their responses are reasonably consistent with the learning objectives and the expectation of the question, and the responses to the question, then I don’t spend a lot of energy writing extreme, extremely detailed, one for one, responses. I don’t offer necessarily a black and white answer, or

my own personal perspective on what the answer to the question should be. I'm not going to write that 16 times for each individual officer.”

Dexter was quick in his explanation that he gave feedback for both summative and formative assessments with the difference being that feedback from a summative assessment included a grade. “...my understanding of the difference between summative and formative was formal feedback vs. just feedback. Formal feedback meaning, they got a grade as part of their overall assessment for a particular lesson or block of instruction that ultimately contributed to their aggregate grade point average for the entire academic year.” He further explained what he felt were differences, “[Summative assessment is] an opportunity for the individual student officer to attempt to execute the task or requirement, and then get specific and deliberate feedback about areas where [they] may have to think about in order to do better the next time...formative is an opportunity too, it's similar but not the same...it's free feedback.”

Ten instructors provided written feedback to students beyond what was required on the CGSC 1002 Assessment Form. An additional formative use of summative assessment was instructors assembling students for discussions of common trends or providing students with written discussions of trends.

Common errors/trends of the group discussed or provided to students

Common errors and or trends of the staff group were provided to students or often discussed by ten instructors during scheduled DTAC class hours or when DTAC classes were not scheduled after the results of common summative assessments were analyzed. This form of feedback is intended to show students what they need to work on to improve and is formative in nature and follows Marzano (2010) who felt that feedback can be given in a group or a one-on-one setting.

Frank explained that he provides feedback on trends to his class as a group, “...and even though, they're officially listed as summative [assessments], I use them when we come back [together after] I grade them and we'll review them [as a group]. When I go back in the class, I review the exam with the students to discuss general trends I saw, you know, both strengths and weaknesses, things I think we need to, stay on track, continue to improve them, or are there areas where I believe we need to work a little bit harder in this particular area.”

Fred said he provides feedback on common errors to the entire class as a group after a summative assessment when he can, “so to the extent that I can, I will try and get in and in front

of the group... do an overview of what the common mistakes [were] ...what I look[ed] for in the assessments, ... if there's a question where everybody missed something, or the vast majority of people missed a concept, or didn't do as well as I would have expected...I go back and do some remedial work on that.”

Caleb took what he had written on student assessment forms and then created a consolidated trends report that he distributed. “I also take all that stuff [on the summative assessment] and I do a trend analysis ... based on what I observed in the assessment, so it's really two-fold: the individual feedback and then I have a little piece that's written that says, here are the staff group trends and here are some recommendations.”

Caleb said he first distributes the assessment forms with individual feedback, and then in a couple of [sessions] he'll explain the staff group trends to his students. “I won't go into each individual thing, but I'll say here are the trends for the group. I generally brief that out, especially during ILE common core, because that's really the baseline where I want to make sure I clear some things up...”

Alex looked for general negative trends within the staff group's summative assessment answers and then assembled his students for discussion in an attempt to increase the group's level of learning, “...regardless of the grade... I want to make sure that [they] understand what right is. So if seven of 16 students didn't get part of the assessment right, then I have to sit and first analyze, OK, was it my failure to communicate, um, no, I think I went through this in pretty good detail, so now it's the students still not getting it, so let's have some, a little more group discussion on that to try to get that last point in.”

Dexter provided his insights to students when he was returned their assessments, “... a technique I use, I don't use it all the time, but when I return the assessments to the group, I will highlight (verbally) particular aspects, or expectations, particular components of a particular question that either they generally all missed, or a group of them generally missed.”

Alfred (Pilot) discussed trends with his students during and after class. “I give a brief summation when I hand the papers out... the good, the bad, here are some of the things that most people missed. Here are the things that everybody was OK with. Then I invite any student to come discuss particulars they have, or questions they have about a comment that I made in a session following the class or whenever we both can get together. But, the biggest thing I provide is ... that out of the 20 questions, you guys have a problem with 'this' concept. If I have time that

day, I go back through that particular concept and make sure everybody [understands because] if 15 out of 20 didn't get it, it's probably my fault.”

Calvin (Pilot) addressed summative assessment trends to his students as a group. “...there were times you see a common type response, and it may not be an accurate response. I'll address that to... get everybody on the same sheet of music of what it should be.”

Allen focused on ‘big’ trends. “I haven't really gone through and shown them trends overall... if I did notice a big trend in the class, then I may bring [it] up as part of going through the answers with them (after returning their papers), and explain, ‘a lot of you had a tendency to answer it this way right here.’ [Then I'll show a correct example and let them know that] this is really what we were looking for, in that manner, right there.”

Archie talked with his students in small groups based on their grade on an assessment. “... After I've graded them based on performance, and I'll pull out three or four students at a time, hand their exams back, and talk in general terms. Here's what I saw on the exam generally from the entire staff group. All of us did this fairly well, all of us answered this question fairly well, and all of us struggled with this concept.”

Ten instructors assembled students for discussions of common trends from the results of summative assessments or because of time limitations, provided students with written discussions of trends. An additional formative use of summative assessment was instructors providing feedback on assessments that students could use in other assignments, including assignments from other departments.

Feedback applicable to other assignments

Eight instructors felt that their feedback given to students on DTAC common summative assessments could be applicable to assignments from other departments or in other non-curriculum related ways. This is contrary to Careless (2006) who found that because instructor's comments were too assignment specific, students could not use provided feedback for other work.

Frank was hopeful that his feedback to students on DTAC summative assessments could be used in courses taught by other teaching departments. “I hope so ...one of the things... we look at [is] how well did you express yourself? You know... if you're writing an essay, how well did you write that essay? ...particularly, have you identified a thesis, and this is how you supported your thesis, which ties back into other classes and lessons where one of the

requirements is to do an argumentative essay. So, I reinforce that. Also...a lot of our assessments get into...how you go about doing critical thinking, and then how you express that critical thinking. So, that skill essentially applies across the board in all our classes here at CGSC. So I think that would be re-emphasized in other courses as well..."

Fred wasn't sure how students used his feedback because he felt that some assessments were written in a way that measured skills not related to what he taught, "... I struggle with that a little bit because we've got, we've got questions that don't lend themselves to essay questions, but the test is having them use an essay format. So a lot of them will struggle to write a five-paragraph theme on a question that doesn't need a five-paragraph theme, and so I definitely struggle with what one of the goals of the course [being] to have them be able to do that. But, it doesn't relate to the material I'm presenting, so... how much do you weight the fact that... they aren't answering it in sort of the format we told them to. Write an essay. Well it ought to have a thesis. If they don't have a thesis, well, you ought to punish them for that, but not having a thesis has no relationship to the skill that we're trying to develop or the thought process we're trying to develop necessarily."

Caleb could tell that his students used his feedback from summative assessments in a couple of ways, some other than the curriculum he was responsible for because of their actions. "One, they've adjusted their study habits, or, [because] they hadn't received feedback [in a particular area], were unaware of that particular weakness [until I told them]. I can see, I know they're applying it because I see consistent improvement in that particular weak area on an assessment as we go along. Thirdly ...I've had a few of them that really take to heart [what I said] and they've developed a more refined self-development plan. Those are examples of how my students, whether it's just an individual, or a collective group, [are] making adjustments."

Allen was not sure if students used his feedback for requirements in other courses. "I would like to think that they can, but I've never really sat down and talked to another instructor to try to figure out did they apply what [I was] talking about to the work that they are doing. I have had discussions with other instructors [like] 'Johnny was doing this, or this is what I saw from Johnny, and this really looks sub-par. Is this the same sort of thing that you're seeing from your classes?' It's kind of a way to get a better gauge of is it something I'm doing that's not really effective in the classroom, is it Johnny having a bad day or a bad week, or is it a trend that

we're seeing overall of, wow, he's really not hitting the mark on anything, it's a bigger problem and we need to all get in here."

Alex thought that because he used feedback from other courses in his lessons, that instructors in other departments are probably doing the same. Alex chose to comment on feedback that he gave students on their writing abilities, feedback that could be used in all CGSOC courses. "I would hope that the same applies conversely... as a tactics instructor, [because I] sit in on DJIMO and leadership classes, I will always try to reference what [students] saw in some other instructor's class as a point. It goes all the way back to ...the C100 block of writing. Take the O400 Stability exam as an example. Even though [students are] only given a paragraph to write, I would say my #1 comment to students this year was, you gave an answer, but I had no clue what your answer was, because you just failed to do the basics of, 'tell me what you're going to talk about, give me two or three main points about that, and then give me a summary conclusion that told me what you told me.' So I think in those types of things I would hope that I have been reinforcing what all the other instructors are saying, in that they are carrying that into [their] lessons.

Dexter felt it depended on the subject area, specifically writing. "...if it's a judgment on their ability to write effectively or communicate, yes, they could apply that. Our current outline of evaluation of writing [is] substance, style, organization, and correctness...the ability to communicate I think is transferrable if it's in fact like a short essay, or a two or three page paper requirement. The aspect of evaluating their ability to communicate and their writing style is transferrable, but in terms of substance or content relevant to the requirement, it's possible that some of that might be tied to the idea of how to think or how to organize.

Jack echoed others that his feedback on written assignments was probably useful on assessments from other departments. "...if it's an essay type requirement, I'll always generally divide it up feedback wise, into general areas. One will be thought organization and writing, and the next one will be specific content. So the former they could definitely use [the] feedback [for] assignments from other directorates.

Calvin (Pilot) described how he graded writing summative assessments, where his comment on clarity could be used by his student' on other department's assignments. "A lot of the words that I use when I grade papers [focus on] greater clarity. I'm [role playing] the staff officer that would be reading this giving me the ability to understand what [the student] is talking

about, [their] visualization. There's been cases where you'll see, an answer that they, the first answer might be very, very short, concise, almost too concise. Where the next time they go around, now you're seeing more in depth analysis of what their answer is. You know, they start out and then they build on it to give the reader better clarity."

Eight instructors provided feedback on DTAC common summative assessments that students could use on other assignments, including assignments from other departments. An additional formative use of summative assessment was providing students examples of good student work.

Use of exemplars

Six instructors provided their students exemplars, or excellent examples of student work, after a common summative assessment. Two chose not to continue the practice, expressing that students should use the initiative to look through available resources to develop exemplars on their own. Garrison and Ehringhaus (2011) felt that "using student work, classroom tests, or exemplars of what is expected helps students understand where they are, where they need to be, and an effective process for getting there." This technique helps students by engaging them in the learning process by establishing and defining quality work together. Chickering and Gamson (1987) promoted the prudent use of model answers, taken either from student's work or specially written by the lecturer, as an excellent form of feedback.

Archie explained that he provided his students exemplars both during lessons and then as handout examples for them to take with them after a discussion of the assessment results. Addressing what he did reviewing summative assessment results, "...the last two teaching cycles... what I'll do is... capture probably the two or three top answers per question, and I'll cut and paste them into a document and hand that out and say, 'here's a couple pretty good answers for each question.' No names, [it's] a pretty good technique that helps."

Allen used exemplars from his own students and he looked for exemplars that fellow team members had identified. "...from talking with the other DTAC instructors on our team, if there's somebody on my team who had somebody in their class who had a really good answer to something, then I'll try to get a copy of that. [I'll] delete the name and show that as an example. ...it's easier to do it that way than if I try to pull something straight from my class, although that's happened too."

Alex didn't provide his students exemplars until the last part of the academic year because of a concern about plagiarism. "...not in common core or AOC (Advanced Operations Course). I will share those things in electives, after we've got to the end of the elective, or they've already left the elective class ...but I think if you did that in core, or in AOC, you may be encouraging some incidental plagiarism [and] I just don't want to go down that path."

Jack stressed that he always obtained student's permission before showing their answers to others as an anonymous exemplar. "I'll put together a two or three slide presentation, and it will encompass without any references to whose results they were, best answers, less effective answers. After I hand back an exam I'll put a couple examples up there and we'll have a discussion about why that's so."

Dexter addressed why he choose not to provide students with exemplars prior to them taking a summative assessment. "I heard a [student] comment that [they] needed to be shown what right looks like, [they] need to be shown the school's solution. There are ...multiple school solutions, there is my way, what I would expect of myself if I was them relevant to the requirement of producing some of the things that we [assign when] we try to build their confidence and competence. And then, what they have to determine is their way... there's plenty of what right looks like available to them if they take the personal initiative to seek that information."

Alfred (Pilot) parallels Dexter's thinking in that although he has used exemplars in the past, he now tries to stay away from them. "I'm pretty hard on my guys. I don't want them to leave here with the idea that [a way] is an accepted way to do things. Or [that] this is what CGSC says we should do. In other words, I don't want them to leave here with "a" way to do it, and think that that's the way it has to be done. Now, if I've got a smart [student] and it's in his [area of expertise], and his answer was completely spot-on, I have in the past grabbed that guy and said, 'why don't you meet with these three or four folks here' or I'll tell those guys, 'Hey, you might want to go see what Ron, or Dave had.' But I try not to give out too many 'a' ways. Or 'the' ways."

Six instructors provided students examples of good student work. An additional formative use of summative assessment was instructors providing individual coaching to students about the results of their assessment.

Individual coaching

Five instructors followed-up common summative assessments with individual coaching sessions. Black & Wiliam (1998b) emphasized that “The dialogue between pupils and students should be thoughtful, reflective, focused to evoke and explore understanding, and conducted so that all pupils have an opportunity to think and to express their ideas” (p. 144). Conversations between instructor and teacher can identify gaps in learning as well as provide the opportunity for instructors to offer specific steps that can be taken to improve student learning.

Allen often initiated follow-on coaching sessions to understand why students might not have done well. “If there’s somebody who had a lot of issues with an assessment or a part of an assessment, I typically, pull them off-side and we’ll talk about that whatever it was that was giving them a lot of problems and [I’ll] try to get a better gauge of, OK, you know, were they just having a head space and timing issue that day, was it a bad day for them when they were filling [the assessment] out, did they truly not understand it. After going through all this [with the student] seeing, hey, here’s what the answer should have been like, does that flash all the bells that need to be flashed in their head to make them realize, OK, now I understand...”

Frank follows-up written comments with individual coaching when needed. “But then also, if there’s a particular problem, I’ll call them in and sit down with them one-on-one and go through the exam and ...we’ll discuss the exam and what I saw, and some areas [where] I think, you know, they could improve upon.”

Fred asks students to see him for coaching when the topic is complex. “If it’s too complex to really talk about in a couple of sentences, anything more than that, then I’ll add, “Please come see me” and work one-on-one when they get a chance to work on it.”

Archie, an active duty instructor, explained in detail the system that he had developed during his several years of teaching. “If I think they have missed the mark on a particular question, or series of particular questions, at the bottom of their exam I’ll say, “Please schedule time to see me”, so individual time. The second thing I do is when I hand the assessments back out, I will, one, again it depends on time because you may be intruding on another instructor’s time as well, and I don’t want to do that, but what I’ll do is I’ll [group] the assessments, after I’ve graded them based on performance, and I’ll pull out three or four students at a time, hand their exams back, and talk in general terms. Here’s what I saw on the exam generally from the entire staff group. All of us did this fairly well, all of us answered this question fairly well, and

all of us struggled with this concept. Overall, the four of you performed above average, or performed in an excellent manner and exhibited a clear understanding of the doctrinal concepts. I have no issues. If you have concerns, you can review your exam and come back to me individually, or send me a note. And then I'll go to the next group of four, which may be four or five folks who've performed to the standard, may have struggled in one or maybe two areas on the exam. I kind of give the same pitch to them, but I'll emphasize the point of, I want you to come back to me individually, not formally, but individually off-line, catch me for five or ten minutes outside of class and just re-explain what you think this concept is, and we'll have dialogue about it. And then the last group are those students who have struggled on the exam, for whatever reasons, and I'll pull them all out and I'll just be frank with them, hey, you guys didn't do so well on the exam. All of you, and that includes me along with you, we need to re-look this exam and just review concepts. So I want an individual meeting with each of you, and then I'd like to set up one, maybe one group meeting after a class for no more than 45 minutes and let's just review all your questions together. You can learn from each other. And I try to couch it in those terms just because it can be a sensitive subject when a student doesn't perform well, especially in this environment, so I couch it in terms of we ALL need to work on this because I feel probably a portion of it is my teaching style which they may not have understood, and my approach which they may not have understood."

Caleb meets individually with students as a follow-on to his written comments. "I'll give them specific feedback, I either hand write it, or whatever the [question] is, or I'll say, I think you're missing the point here, it's not in accordance with doctrine, or it may be another question that says, did you consider more integration or this aspect of transitions or integrations, and I do that individually [in person]."

Caleb continued, explaining how his coaching included his expectations of students. "I am always available. What I've tried to do is to take maximum advantage of the classroom time and where I've seen people struggle; I've worked [with] them, [on] a one-on-one basis... However, I have placed that back on the officer, and said I see [your problem area]. I'm available. Let's make a plan, and unless it's going to cause them some angst in terms of major grade issues and failure, I really leave it up to them to come see me. Now, if it's a C or a U, absolutely not, I'm really forceful but they're professionals. If they want help I expect them to come seek me out. It isn't like I, uh, now obviously the remediation that's another issue. I do

help remediate. Uh, but I've done that in terms of the time I've been here teaching both military and um, civilian, I've really only had to do that with 2 students, and they weren't US officers. It was a language issue.

Calvin (Pilot Interview), unlike other instructors, leaves it to the student to initiate contact about meeting for coaching, "...because we don't have the time ...I don't take the time in class to say, "okay, here's question one", because there are so many different responses to it. I'd like them to have them come back to me if they have questions about something, to clarify what I wrote. Or if they disagree with what I wrote, then I give them the opportunity to come to me and address it with me."

Five instructors provided individual coaching to students about the results of their assessments. An additional discussion relative to the ability of feedback to be formative was the lack of timeliness of feedback to students after a summative assessment.

Timeliness of feedback

Five instructors felt that the lack of timeliness of feedback caused by the way curriculum is scheduled affected student learning in a negative way. Light and Cox (2001) felt that the scheduling of assessments is usually too late to allow for timely feedback, denying students a sense of their own progression. Feedback that is not timely contributes little to student learning (Chickering & Gamson, 1987).

Fred felt that the timeliness of the feedback he provided to his students was a problem out of his control. "I struggle with this one because generally the way the course lays out, by the time we can get around to grading it, they're off into another block [of instruction], and so in order to get back to the student to give them feedback, you have to cut into some other instructor's hours to get in front of the students again. So to do it in your own hours you have to wait maybe a month before you get [the opportunity], and by then the power of that feedback is diminished to some degree."

Alfred (Pilot) defined 'timely' in terms of his student's being able to apply his feedback in future blocks of instruction, not the current one. "I think my feedback is timely. Now, whether or not they can process it and whether or not it applies to future instruction, I guess really depends on where we are in the curriculum. I think it's always a lesson learned they put in their kit bag, and they can keep it in mind for later concept if it happens to them in a unit ...post ILE. But, you know if it happens to you in a class early on, for example, if a concept's not understood

in C500, and feedbacks given, I think they have the chance to apply that later in the AOC. So, yeah, I think it's timely."

Calvin (Pilot) addressed the ten working days that CGSOC instructors are authorized for grading assessments and returning them to students. "Yeah, sometimes it's longer, and it depends on what we're doing. Ten days is 10 days. That's a lot of days. However, when you're in the classroom every day it's hard to [make 10 days], especially when you have [an assessment] where you have to sit down and physically grade it, and it's a paper that has multiple questions on it, where you have to read and analyze what they're talking about. And, you've got to prep for the next day of class, and you're going into the classroom. Hours in the day ...with everything going on, a lot of it you either take it home, you do extra time. But you try to hit the 10 day line. There's been times when I've been over 10 days. There's been times when I've been under 10 days."

Frank put the problem in the larger CGSOC context. "...I realize the problem that we have in timing, particularly that our curriculum has to tie into what's going on at the TASS sites and the satellite campuses, and that sometimes drives our timing and affects what we have to do. Also, we have the other departments, you know, what's going on student-wise with all the different assessments. So, we have to balance it out so you're not hitting them with everything at the same time, so that they have some time to put some quality effort and reflection into, to their assessments and into their work."

Allen echoed Frank's understanding of the problem. "The assessments, their timing, those sorts of things, I think the timing is, is pretty much based on, everybody would like to be able to spread things out a little bit more, and those sorts of things there, but we don't have a lot of time in the classroom where everything's compressed anyway, so it kind of goes where it goes. I think it's really how to those assessments and their timing for when they're due, when they're assigned line up with all the other departments in which I think they do a less than adequate job across the college spacing out so that the students aren't overwhelmed with assessments, which happens too frequently."

"I think an important thing is getting them the feedback, back to them as quickly as possible, because they're on such a quick pace academically that you go beyond a couple weeks it's, ok great, yeah, whatever, I'm onto the next three sets of assessments that I'm doing... I think with a lot of them, it probably takes more effort to try to remind them of, OK, here's where

you had issues with the previous assessment, here was the feedback that you received on it, we've got another block of stuff that's coming up, take that all into account as you go through and do your assessment for your next block here.”

Five participants discussed how the lack of timeliness of feedback to students after a summative assessment affected its usefulness. An additional formative use of summative assessment was instructors providing references for students to be used for further learning.

References provided for further learning

Four instructors provided references for students to read for a better understanding of topics they did not understand on a common summative assessment. This follows Black and Wiliam's (1998b) recommendations, “Feedback on tests, seatwork, and homework should give each pupil guidance on how to improve, and each pupil must be given help and an opportunity to work on the improvement” (p. 144).

Archie was very clear in his approach, “I'll [write] for example, your comments or the way you structured your commander's intent, didn't necessarily hit the mark, go back and review ADRP 3-0 and then I'll site the page and paragraphs. [I'll] ask them to review it, and then, I may add the note, come back to me and tell me what you understand, though that doctrine says how you would re-write your commander's intent to fit with what doctrine says, to frame it in a better way.”

Allen typically provided written comments when students cited an incorrect or outdated reference or no none at all. He added that he would look at assessments ahead of time, and let students know that, “there's a whole, wide range of this, of references you can use, I'd like you to narrow your search to these one or two, or three, whatever it is, so [you] don't go too far off on a tangent.”

Alex explained a technique that he had developed over time. “No one ever told me to do this, my general technique is [to use] all the pro words that are associated with the universal and actual standards; clarity, significance, relevance, precision, and action to comment positively or negatively on a student's work.” Then he recalled a specific event where students had used outdated Army doctrine, “Some students were using terms and symbols that were in the old FM 1-02, not the new ADRP 1-02 [and I provided comments telling them where they should be looking].” Speaking rhetorically, Alex continued, “So, to me what's the whole purpose of giving

the student any feedback? It's to improve their learning. When the assessment is done, the grading is done, but the learning should not have ended. So, if I don't give them where to go find what right looks like, then I haven't done my job."

Jack was very succinct about providing students references for further learning, "Yes, always." Jack and two other instructors provided students references, thinking it part of their responsibilities.

Descriptive feedback is a characteristic of formative assessment, not summative assessment. Participants chose to use provide students feedback in many different ways, with the intent of improving student learning the essence of formative use of common summative assessments. The modification of teaching and learning activities also normally associated with formative assessment, was used by participant instructors after summative assessments.

Theme Two: Modification of Teaching and Learning Activities

"Formative assessment is a planned process in which assessment-based evidence of students' status is used by teachers to adjust their ongoing instructional procedures or by students to adjust their current learning tactics" (Popham, 2008).

Instructor comments about their modification of teaching and learning activities because of evidence from common summative assessments resulted in the four findings below. This includes a finding of recommendations to modify future common summative assessment instruments which would have the effect of modifying future teaching activities. The numbers in parenthesis indicate how many different instructors out of 12 participants (10 primaries plus two pilots) addressed the topic/subtheme:

- Recommended adjustments to the common summative assessment instrument after the assessment was given (9)
- Adjustments to learning activities or teaching style after the assessment was given (7)
- Adjusting learning activities or teaching style before the assessment was given (6)
- Reflection by instructors about the need to modify teaching or learning activities after the assessment was given (5)

These findings align with Black, Harrison, Lee, Marshall and Wiliam's (2003) definition of formative assessment:

An assessment activity can help learning if it provides information to be used as feedback by teachers and by their pupils in assessing themselves and each other to modify the teaching and learning activities in which they are engaged. Such assessment becomes 'formative assessment' when the evidence is actually used to adapt teaching work to meet learning needs (Black et al., 2003, p.2).

Recommended adjustments to the common summative assessment instrument after the assessment was given

Common assessments are formative or summative assessments that are created collaboratively by teams of teachers who teach the same class or grade level. Nine instructors discussed recommendations they made to curriculum developers after reviewing the results of their student's common summative assessments. The level of instructor effort to provide recommendations to adjust their common assessments supports Stiggins and DuFour's (2009) argument that common assessments for formative purposes, or "assessments created collaboratively by teams of teachers who teach the same course" (p.640) are a powerful tool for creating effective assessments.

Frank explained the DTAC process for recommending changes including how his teaching team operated internally. "... after every block, DTAC conducts an AAR (after action review). And what we do before the AAR, we'll have a team meeting where we'll discuss how we thought the lesson went, where we saw the strengths and weaknesses and things that need to be changed. So you know, we're sort of comment unrestrained across the board on how each one of us saw the lesson, and then whoever goes to the [DTAC formal] AAR or submits the AAR comments [to the curriculum developer] will submit them from a team perspective."

Frank continued, "... not every comment that you have is going to elicit a change, but I think the comments that we do send in people look at them seriously and evaluate them within, across the board with what some of the other comments are. But the comments, [in] my experience ...can be ...radically different from team to team, so you sort of have to figure out, what the medium is, and adjust from there."

Later in the interview, Frank addressed specifically the creation of common assessments in DTAC. "...I think it's pretty much a group effort. I think ...progressively, it starts with maybe one or two guys, but as they come up with a prototype, they send it to other people throughout DTAC, so that there's different eyes on the assessment that, you know, make sure it's hitting the key points, that it's a valid assessment, and, you know it has buy in and support from the faculty."

Fred described how the process his team used to gather input for AAR (After Action Review) comments that will be forwarded to the Block Author. "...after a block we'll get a, 'hey the AAR is here and your team needs to send somebody,' and we'll figure out who on the team is going to go... the guy who's going to represent the team. And then, everybody else will talk about it over lunch or send e-mails and say this is what we liked or didn't like about it."

He continued, "I'm frustrated with, I don't see a clear link between the questions we ask on the exam and what we think the most important concepts of the course are. For example, the concepts of Defeat and Stability Mechanisms, in 3-0. Defeat and Stability Mechanisms aren't even important enough to have their own title. There's not even a sub-paragraph bolded title for those two concepts. They're talked about vaguely in a sub-paragraph. Yet, one of the questions on the test, I think is worth 30%, or 40% of that test, is an essay about Stability and Defeat Mechanisms. If it's not one of the most important concepts in 3-0, then why is it one of the most important concepts on the test? If I were a student and I left the course, I would think to myself, the DTAC folks think that this concept is really, really important, but the [doctrine] writers clearly don't. So, as a relatively new instructor, I don't understand how that particular question became so important relative to things like The Decisive Point, or, I mean any number of other concepts that are out there. So, it's not really clear to me who picks which questions get to be on the test. Some of them are kind of handed down, and I don't know that we clearly link, what are the five things we want the student to leave here really being able to do?"

Archie explained how recommended changes to assessments were gathered on his teaching team. "We have our own kind of internal team process, in our case on our team, [name deleted], or [name deleted], collects our feedback, and then when the block author asks for comments on the block and each lesson, we [provide] our DTAC team [deleted] team's consolidated feedback. So, you know, we'll talk [about] what are the key points we want to send forward, and we agree on those points, and then we have one guy forward them for us. So [the

block author] isn't getting 16 individual, or 4 individual answers from team [deleted], he's getting a unified feedback approach there."

Caleb explained how he coordinates recommended adjustments to the curriculum, "... on our team we do it a couple of ways. I'm the lead instructor on our team, so in terms of DTAC curriculum, [team members] give me feedback and I consolidate it and send it to our curriculum developers. But I also CC: my team leader so he's aware. I have also [shared feedback] with my department of joint interagency and multinational operations teaching partner (DJIMO), because a lot of their concepts directly translate to what we do. Many teams do this. They've designated a lead person for each department to be the conduit for feedback... back through the team leader to the appropriate curriculum developer. That's good in terms of giving feedback from a team perspective, however, if not all the teams [are] ...consistent, some of the adjustments that are recommended may be made in a vacuum and it may further disintegrate some of the concepts that we're trying to get to which is exacerbated by a stand-alone, non-integrated assessment."

Allen chose a specific assessment to explain his concerns with how assessments were designed. "I think it was the O300 some of the COA Sketch and Statement is one that really stands out as we all sat down first of all and looked at the rubric that came out for it, and then got through with our initial look at it, and say, ok well here's some things that we can see already that are going to cause problems, so what do we need to do across the board within our teaching team to try get some commonality on that. ...And then at the end of the block, going back to them and saying, OK, here's you know, the 28 or so extra issues that we found with this particular assessment and kind of going through and explaining the reason why for all that piece of it there. Also, talking within the teaching team, the piece of trying to make sure that one of us isn't essentially being the candy man... OK, I found an extra 40 gigs or whatever that they could have picked out on that particular assessment. Well, you're really going outside the bounds on what you're giving them on this stuff here." The terms, "candy man," or "Santa Claus" are used by instructors to describe other instructors who are easy graders.

Like the other instructors, he expressed confidence that curriculum developers were trying to do the right things. "I definitely feel like most of the course authors are trying really hard to get feedback from all the instructors. I think after this first year I've got a better idea of how that feedback process works and that yeah, what I say can be valued."

Alex explained “at the end of a block, when they [ask] for AAR comments, I’ll send them [to the curriculum developers]. ...last year we consolidated some comments as a team and sent to [name removed]. As a team we didn’t really have any consolidated team comments this year.”

Jack explained, “I don’t have a process. I usually will comment before the exam is made. Usually the course author will pitch it to either DTAC as a whole or [omitted]. I’ll provide input then and then of course, afterwards, but in terms of an actual process, no.”

Arlen described how his team included feedback from his students, “On the team I was on we would solicit comments from the students. I would save an hour at the end of a block of instruction and ...we’d get with the students and say, ‘what did you like about this assessment?’ [We] had other ways of putting in feedback, but I think giving some feedback to the instructors is important, and capturing that, and then we’ll roll it up as a team to get it back to the lesson author ...we like this, we didn’t like this, and that was helpful. And then, as I would say, as an instructor, I didn’t do that enough. Feedback is good, no feedback is bad, and I think we did an OK job getting some feedback at the end of a lesson.”

Hank described what he saw as an assessment feedback loop, “...at the end of the block we always have a discussion in class about how it went, not just in the class, but [students] give feedback as to what [they] thought of that assessment and generally we’ll take that, and then we’ll feed it back to the curriculum developer, who’s tied into the assessment for the block. I look at what they wrote versus what question was asked, I look at was there some kind of misunderstanding, like the answers are a little far off. Or was it really testing what we should have out of this block? Was this a, look it up in the book, which doesn’t require any learning, look at the table of contents, or was this one that required them to really think through and analyze and come up with a solution? So, I think we do have a feedback loop on the assessments whether it was effective or not, not only from the student, it kind of goes from the student to the faculty member, who will then flow it back to the curriculum developer.”

Nine instructors provided feedback to curriculum developers after a summative common assessment was given. Recommendations acted upon and included in future curriculum will cause changes to teaching and learning activities, including summative assessments scheduled to be issued during the same block of instruction. Instructors also used information gathered from assessments to make adjustments to their teaching style or classroom activities.

Adjustments to learning activities or teaching style after the assessment is given

Seven instructors discussed how they made adjustments to their teaching style or activities after reviewing the results of their student's common summative assessments. The adjustments were not immediate, a characteristic some scholars say defines formative assessment, but rather occur during the Academic Year, while there is still time to take action.

Fred felt that assessments were an indicator of his teaching style. "I definitely try to go back ...because to some degree I think the summative assessment is as much a review of how well I've presented the material. So I definitely go back and look, across the sixteen students, are there any trends that I can see across all the answers that would indicate that I'd either done particularly well on a concept, or that they don't misunderstand something I've said, or weighted things wrong, or, whatever the perception is. So, I look to see if I need to go correct and re-teach anything."

Fred felt that assessments were almost more helpful to him than the students. "The Azeri course of action, for example, based on the common errors that people were making, I definitely made an attempt this year to find places and examples, you know, when you're laying out a course of action, or when you're teaching in GNI, hey, here's a common mistake, your peers don't understand how boundaries work. And this is because it kind of, there's a generational gap between how we experience, and what we think is sort of standard and what they think is standard. You can't really see that until you've seen a bunch of them in practice and understand that in order to get us closer together; there are some things I'm taking for granted that they know that they just don't know. So, I think those assessments are almost more helpful for me than it is for the students."

Later in the interview, Fred further explained, "I'll try and carve out sometime in the next block to go back and say, hey, here's a concept that I don't think I did well enough to teach because most of you didn't grasp that question. In three years I've had two questions where I don't feel like the group had got it and I had to go back and do some remedial work on that. If everyone's generally getting it and performing about where I expected, then I usually won't cut into to do an AAR of the assessment. I personally think we ought to have some time built into the curriculum that's our hours, our face-time dedicated to the review of the material. But right now you have to carve out of some other topic to do that."

Archie related his teaching style and techniques have progressed over time. “How has grading and review of summative assessments adjusted, or impacted, [my] teaching style, teaching methodology? I mean, my first couple years here they did not impact at all, but that’s because I didn’t know what I didn’t know, in terms of teaching. It took me awhile to figure out that that’s a variable that should be included in my assessment of whether or not students are grasping doctrinal concepts and applying them effectively or critically, when I teach a particular block, or particular lesson. So the last two years what I’ve done is I’ve taken, and this goes back to our own internal team [deleted] DTAC huddles after each assessment. I’ve taken what questions did we struggle on as a group, and then go back, take that and kind of reverse engineer it, how did I teach that particular topic, or those topics that apply to this question, and where can I adjust. The next time I teach this, how am I going to do it differently? So, that’s one way.”

He contrasted teaching electives at the end of the academic year with teaching during the core, or foundations block, the focus of this study. “I think the second adjustment, to open the aperture a little bit broader, is [that] how we teach electives is drastically different sometimes than how we teach the core or the AOC curriculum. It gives you a little bit more latitude to refine your craft in terms of teaching, and you can observe some different teaching styles when you’re partnered with different guys, and I think that’s helped kind of change my approach to how I teach core lessons, in a block, and then how I apply, those lessons learned to teaching styles in the core blocks, if that makes sense to you.”

Archie recognized that he became more directive in his teaching style when teaching the kind of concepts that appear on assessments. “Yeah, adjust fire, I call it adjust fire. I don’t want to say change. Deliberate change, and part of that’s a little bit nuanced, you’ve got to understand where your students are, so I kind of ask them leading questions, to make sure I’m kind of gauging they clearly understand this doctrinal concept A, but they’re struggling with doctrinal concept B and C. They clearly understand mission command as a philosophy. They are struggling with mission command as the science of control, and how mission command and science of control fit together. So you’ve got to gauge that and I’ll adjust the lesson plan or, adjust fire to incorporate that into whatever dialogue we’re having for the day.” I’ve only noticed this over the last year or two, while I’m teaching the concepts, I want to pull, draw in student dialogue as much as possible, and that may take some additional time, and so there’s a good dialogue back and forth and the students are learning from each other and not just hearing me

talk, but if it gets to a point where I'm concerned that they don't necessarily grasp, a particular concept, I become a little bit more directive in my teaching style and, I kind of become the focus of the dialogue in explaining the concept to them in a little bit more depth or detail, at least to give them a clearer picture or a clearer understanding of what the concept is and how it relates doctrinally to other concepts or to tactics, or to operations."

Caleb's answers showed how he adjusted throughout the curriculum, taking advantage of knowing that he will have the same students in class during later classes and will be able to readdresses concepts iteratively. "Initially, based on the assessments, if it's [during] ILE common core and I see a weakness, and it's a consistent trend, then I often will re-adjust. When I'm later on in the curriculum, and I can't do it all, but maybe this one particular thing that maybe they weren't quite getting, I will adjust part of my teaching plan in AOC and later lessons where I can hit that again and see if it was that maybe they didn't receive or maybe they misunderstood, or maybe they got knowledge now that they didn't have before either from me or another class, and I may adjust my teaching plan in the near term for those subsequent lessons. That's one way that I use that. The other thing is, I solicit feedback from my students and I have little survey things that I do, periodically. Do you understand the assessment, and do you understand my feedback on the assessment because maybe my feedback to them is the problem, maybe it's not clear, maybe it's not, appropriate or whatever, but I request direct student feedback. So that's another way I use it. As far as adjusting for the next year, absolutely. And ...maybe it's an area that, maybe I focused on this war fighting function too much and maybe I didn't look broader so maybe on my feedback I need to, instead of concentrating on these three, maybe I need to expand that. So I may adjust how I approach in not only common core, but a particular lesson. But the challenge there is, I mean and it's more self-awareness as you as a teacher in terms of my assessment and how I'm teaching, but the other piece is you've got to be guarded against swinging the pendulum too far because every staff group every year is different. Their knowledge base is different, and so you're constantly having to do an in-stride adjustment anyway if they're not receiving what you're trying to articulate. So those are just examples of how I use that information through assessments, to either adjust my teaching plan or to continue to re-connect the dots in the current curriculum."

His continuation shows how instructors seek out ways to integrate lessons from other departments within their teaching teams. "I generally will sit-in, and other guys sit-in on my

class so that we can integrate and see where those weaknesses are in the group, and that's a good technique in terms of seeing where we may need to collectively adjust our teaching plan. But I've not had an issue going in, from my experience, going in another block, maybe for 10 minutes or during the tail end of SGA time or whatever it is, kind of going over that stuff. Perhaps a better way, or a more effective way, organizationally, would be to have let's say three assessments in one week and you've got 10 days to [grade each] each. Maybe another way is to at the beginning, you know, hey, let's review all these assessments we did. They can do it all at once. Now, there's, there's pros and cons, you know, with that, but there's different ways to do that but I haven't experienced any challenges on my own team by doing that."

Caleb concluded by mentioning self-awareness and having to continually adjust teaching styles/activities. "...the challenge there is, I mean and it's more self-awareness as you as a teacher in terms of my assessment and how I'm teaching, but the other piece is, you've got to be guarded against swinging the pendulum too far because every staff group every year is different. Their knowledge base is different, and so you're constantly having to do an in-stride adjustment anyway if they're not receiving what you're trying to articulate. So those are just examples of what, some examples of how I use that information through assessments, to either adjust my teaching plan or to continue to re-connect the dots in the current curriculum."

Allen spent time assessing his teaching performance and how he could adjust to readdress topics he felt students weren't clear on. "I guess what I do is probably something similar to what a lot of other folks do, which is, as I go through and try to figure out the trends of what they did really well, in some cases it's surprising to me and, wow, they really hit that and I didn't think I hit that, you know, as well as I did in the class. Or, one that they did really poorly on, to try to figure out, okay, was it an issue with how I presented it, how much emphasis I put on it, the examples I used in class, those sorts of things there, and then as I make out my AAR notes for the different lessons, trying to figure out OK, what can I do differently down the road, for the next time when I pitch this particular class, number one, and then number two, for the group that I've got now, OK, when else do we have this, or a similar set of points that are coming up to be talked about in a later lesson block, and how do I need to adjust my teaching method or what I'm doing for PE's, or to try to do a better job of re-emphasizing those sorts of things. On the Defeat and Stability Mechanisms, the way our curriculum is set up you're hitting it in the C500 block but you're still going to come back and hit it again in O300, O400, so those

are things there. Probably something else that I could do would be getting a better feel for, example, what DJIMO does in some of their blocks. And, how can I do a better job of weaving things together with them on some of those things? Some of the more experienced instructors are really good about those types of things.”

Allen continued, “...there’re some things that I’ve got to do to get better as an instructor of, incorporating the assessments into how I pitch the lessons, I think a lot of that comes with practice, and doing a better job of organizing my time as an instructor, those sorts of things. I think I’ve also got to figure out more tools in the kit bag of different ways essentially to pitch things. You know, whether it’s PEs or those sorts of things, so students get more exposure to those sorts of things, and then figure out OK, how do I get them more iterations of the things that they need to apply and that are on the assessments, those sorts of things.”

Alex, like the other participants, constantly evaluated his own teaching ability when determining how to adjust future sessions to cover topics that need to be readdressed. “I think the second thing that I look at, were there common trends within the students, of everyone struggled on this particular question. That’s where I think you have to go back and do the self-analysis of, if most of the students struggled on a particular question, was it the way the question was worded? From an instructional, an assessment design standpoint is that the problem? If the question is very straight forward and very clear, then about the only conclusion I have left is, I didn’t do a good job discussing that, and maybe it goes back to the difference between a first year and second year instructor or, I chose to stick with exactly what the lesson plan was, and maybe the lesson plan didn’t really touch that, so I sort of failed my students there. So there’s that balance.”

Arlen thought more about adjusting his teaching during the next academic year than the current one. “...I think if the students really miss a question and it’s a consistent theme throughout, and they all missed question three, or did not do well on it, I’ll leave a note in my file for that class as an AAR common aid that this wasn’t covered very well, or maybe look to, [telling the] course author to look at revising that question for the next year. I noticed after reading some of the questions in the O400 block that ...some of the questions were really redundant. You know, two of the questions asked the same thing just in different ways and, you know, until I looked at the answers from the students, I didn’t realize that, but definitely, I think, based on what they do, you know, I’ll make a note, if I saw where an area they missed, maybe I

would change the way I instruct next year. Highlight that or to change the assessment or the questions on the exam.”

Frank also looked towards the next academic year to make adjustments. “When I go through it I make notes to myself. Sort of like my own AAR. The next year when I do this these are the things I may want to, you know, focus on a little bit closer, or maybe adjust the way I go about doing it. “One thing that I have found, one year, one class may be weak in one area and stronger in another, and then the next year it could be the exact opposite. But I do try to, OK, these are the lessons learned, these are some things I think I could do better, or based on what happened this year, this is how I would change it. So, I do try to incorporate that into the, the following year.”

Later during the interview Frank continued, “...that’s why I say even though they’re officially listed as summative, I use them when I grade them and will review them, and then for follow-up lessons I identify a trend, [for example] across the board I thought we were weak in this area. Then I’ll go back and re-emphasize that, and maybe re-adjust the way I’ve gone about teaching it. You know maybe I’ve done something wrong, something wasn’t quite clear. We’ve gotten confused and we need to adjust our track.”

Seven instructors described how they adjusted teaching or learning activities after student assessments were given. Modification of teaching and learning activities can also occur before an assessment is given to students.

Adjusting learning activities or teaching style before the assessment is given

Prior to issuing a common summative assessment to their students, six instructors described how they planned to adjust their teaching activities after reviewing the assessment instrument. Most summative assessment were take-home, allowing students several days to one week to complete.

Allen “...that’s definitely one of the first things I’ll do is, in addition to pulling up what the lesson plan is, okay, here’s what the assessments are. Because that tells me, at a minimum I have to make sure I cover those points in sufficient detail to go on with this. An example would be, say, Defeat and Stability Mechanisms. While maybe that’s not something that I personally would think of as, you know, a super key point to really hit, on one of the tests that was

something that was emphasized was [so] OK, I really need to make sure I pay attention to this, especially since it can be confusing.”

Later during the interview, Allen continued, “... if I know it’s going to be on an assessment, then I’ve got to figure out ways to not only during the GNI (generalize new information) hit it with them, but also come back and show multiple different ways that it can actually be applied. And then, as they’re doing the PEs, try to tweak the PEs to the point that they’re actually applying those different principles, or things that they’re going to be asked. I don’t want [it] to be the first time they’re actually going to use it is during an assessment. I want to at least have done it a couple times in the class.”

Alex explained, “I probably do change my teaching style somewhat if it’s on an assessment and I say that because, I want to make sure, to the degree possible, that we have worked through what right looks like in the learning environment of the classroom. So if I know that they’re going to be evaluated on something, I want to make sure that we work through it. And not all of our lessons have a PE, practical exercise associated with it. Most do, but I think that you sort of owe that. At the same time, it’s never giving away the answers. My expectation is we have had good classroom discussion, we have had a good in-classroom practical exercise, we all know what right looks like. When that general topic comes back up in an assessment, then I expect a higher level of performance from my students and there’s no expectation, you know, there’s no latitude for them of, well, we really didn’t talk about this in class, so I can give you a very general answer.”

Hank explained how he adjusts future PEs (Practical Exercises) after looking at an upcoming assessment. “I will make sure that I will bring out points in the class that I think is needed for the summative assessment. Then I will craft the practical exercises to also highlight the points that are going to be asked of them in the summative assessment. So yes I do adjust it. So if I don’t think that the practical exercises [will] highlight it enough, then I will bring up, bring out teaching points, when I’m doing my part of the presentation, to ensure that at least we’ve had the discussion in class, so when they get to the summative assessment, it’s not something that’s totally new to them.”

Hank continued, “I don’t think any of my classes have gone the same way to tell you the truth. So, yeah I’ll look at the notes, and sometimes, even the ones that I said went really well, well I try it the next year and it won’t go as well. So, I’m always careful. I think almost every

class I will go through the material, I will brain storm how I think the flow of the class will go and I make sure that I've got the points that I want highlighted, and every one of them I feel like I'm, almost crafting from scratch. It's doesn't always go the same way every time, but I always do look to try and improve it every time, and sometimes it works, and sometimes it doesn't."

Caleb was very detailed in how he planned to teach after looking at upcoming assessments, "... [I use] couple of techniques, and it really depends on the subject matter. As part of my preparation, what I generally do is I'll look at the assessment, what the assessments are for based on the assessment plan. If it's a course of action sketch and statement or a commander's intent, or maybe it's some kind of exam that's midway or whatever, I try to look at that, and it should be nested and linked to the learning objective. However, sometimes it's not clear in the lesson plan. So generally in my preparation I will start with that. 'What is the assessment, a formative or summative?' I will balance that against the specific lesson to make sure if there is not a direct linkage I understand where I need to fill that gap in either my explanation, during the actual conduct of the assessment, or, I'm going to have to adjust my approach and my teaching plan to lesson plan. For example, ...if we're doing discussion of doctrinal concepts, my expectation is they'll have done the reading and the doctrinal readings based on our department's requirements, and I generally will start with some kind of opening question, and I try to get them to visualize on the [white]board. So if it's doctrinal, I'm trying to get them not to regurgitate the doctrine, but trying to get them to show me the linkage from this concept to this concept if it's doctrinal based. That's one technique. Another technique is to help them better understand ...where there's an Army specific concept linked to a joint concept, I will start by a review. Let's review, for example, C304, which is the Army Organization and Capabilities. Really talking higher echelon armies, service component command and all that kind of stuff and asked them, in terms of capabilities, let's review, now that we've seen this big army piece as part of the joint force, we're ...going to get them to connect the dots. And what I found technique wise, is that helps them at least get a broader context instead of this lesson, this lesson, then this lesson, before even start going into for example, detailed planning."

Later during the interview, Caleb continued, "I've been on teams that do it different ways. I generally always will discuss the rubric with [students] when I'm handing out the requirement. Go over the instructions, make sure if there's any kind of pen and ink [updates] we've got to do, they write it down, and then I'll go over the rubric. And then I'll even amplify

that it terms of some of our rubrics have a one to five, [Likert scale] and it kind of has some broad language, but I try to articulate it in usable words for them. But even then it's tough, in terms of some of the assessments. For example, Commander's Planning and Guidance - in our doctrine, it says there is no set format, however, here's some things to think about based on war fighting functions in this table but, Commander's Planning Guidance is art. And it's hard to judge art based on a scientific table. And so that rubric does not lend itself to an objective evaluation based on the rubric. So, I try to fill that in, but generally the rubric, in terms of the assessments, and explain it to them. But that's still a continuing challenge from my perspective."

Arlen focused on ensuring that he covered everything on the assessment in class prior to issuing the assessment, "...I look at the students, and for any of the assessments that we have, try and look what's in there, and then during the class time, and the discussion, try and address those points, so at least they have a reference, a starting point for those items, and then an area to focus on. So I'll look at the exam questions and try and incorporate it into the GNI (Generalize New Instruction) instruction that we're doing and try and make sure those points at least come out and we have some discussion on them before the test is given."

Frank described how he took the assessment himself to ensure he understood it before teaching the content covered in class. "I review the exam at the beginning of the block to make sure I understand the exam, and also I work through portions of the exam myself, you know, to see how I would do on it. But the key thing is to make sure I understand what the exam is, what the exam is going to address, and then just to make sure that I don't see any disconnect between what we're going to cover in class and what the exam is going to focus on." Frank was one of six instructors who reviewed the assessment instrument prior to issuing it to their students and planned to adjust their teaching activities because of their review.

Six instructors described how they adjusted teaching styles and or classroom activities prior to issuing students a summative assessment. Several instructors described reflecting specifically about how they were going to teach future classes the assessment was given.

Reflection by instructors about the need to modify teaching or learning activities after the assessment is given

Five instructors felt they received useful feedback from common summative assessments given to students that caused them to reflect on how they taught certain topics and if they needed to modify their teaching or learning activities.

Hank felt that having taught his classes and looked at assessment results that he was now prepared to teach. “When I look at the results of their summative assessments, I say, wow, did I not teach this right? If it’s not good I think most of us are our worst critic... I always think, did I teach this right, is there a different technique that I should have used that would have got these points across better. And so I always do a little bit of a self-analysis, and then I think sometimes you’re ready to teach after you’re done. You know, you say, OK, I can do this a lot better if could just reset the clock and do it all over again. And so, that’s not always a great feeling, but you’re always better prepared to teach it after you get done. And so I think I do use what they give me, as their summative assessments, as feedback to me as to how well I taught, or didn’t teach.”

Hank continued, describing his feelings on the usefulness of feedback to faculty. “I do believe that the assessment results are more important to the faculty. I believe that ultimately you want your students to do great on all the tests. And if they don’t do well on the assessment, I think it’s great feedback to the faculty as to how the class went. Did you create the right environment, did you get them into the learning so when they do the assessment they do well. It’s a double edged sword here, you can obviously, you know, teach them for the assessment and it’s great, but that’s not really what I’m talking about, but, I’m talking about, are they into the subject, are they learning, and when we get to a summative assessment that really gets into, a higher cognitive level, they do well because they can put all these concepts together. Well that’s great feedback to the faculty and that’s very useful, and if they don’t do well I think that’s great feedback for the faculty that I really need to look at, how I constructed this class.”

Jack disagreed with Hank, explaining how assessments are more important to students than faculty, “...when I see a student or a couple of students who are generally very good students in terms of their understanding and work ethic [and] they don’t do well, I always have to look at myself first, and review, what it is that I emphasized or didn’t, in class, and take that into consideration. ...I don’t think that [assessments] help instructors at all unless they are simple things that are easy to grade, and that’s not an effective measure of an assessment. I think the better [assessments] help the student far more than the instructor.

Alex prefers to look at written comments from students to help in assessing his teaching effectiveness, “... and a third thing that I would do which is very general, is that at the end of the

year, I hand them out a survey, an assessment of myself as an instructor and get that feedback to look at and do self-assessment.”

Fred felt that assessment results reflected his teaching as well as how much students learned, “...because to some degree I think the summative assessment is as much a review of how well I’ve presented the material. So I definitely go back and look, across the sixteen students, are there any trends that I can see across all the answers that would indicate that I’d either done particularly well on a concept, or that they don’t misunderstood something I’ve said, or weighted things wrong, or, whatever the perception is. So, I look to see if I need to go correct and re-teach anything.”

Dexter was perhaps the most reflective of his teaching style. “The first thing I reflect on is myself and my opinion. A couple different things happen. In the interest of the students, I am critical in my evaluation of the assessment itself. Whether or not it was well organized, that it was clearly communicated in simple written form, separating myself from that responsibility, if I picked it up and read it, could I understand in fact what I’m supposed to do. So that’s an aspect of consideration that I include in some of my personal evaluation of their responses to the requirement. Secondly, I reflect on whether or not I did my job... based on an assumption that one, if I have been successful and have provided, provided the opportunity for them to study, learn, explore, and discuss, the elements of the courseware or the lesson framed by the learning objectives. Did I in fact set the conditions for them to be able and prepared to be thoughtful in their response to the question? So, I’m reflecting on myself at the same time based on how I see their responses. If the bell curve is wide, and I’m seeing a lack of performance in the majority of the group, then that really challenges me to go back to that thing I just mentioned and that’s whether or not, did I set them up for success or failure in my own personal, you know instruction?”

Instructors modified their teaching style or learning activities either before or after students took common summative assessments, reflected on their teaching styles and made recommendations to improve future common summative assessments, all examples of the formative use summative assessment. The following section transitions from a formative focus to similar practices instructors used as a result of common summative assessments and addresses findings in the areas of teaching and grading techniques.

Research Question Two

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

Theme Three: Preparation of Students for Upcoming Summative Assessments

"When teachers match their teaching to what they expect to appear on state tests of this sort... students are likely to experience far more facts and routines than conceptual understanding and problem-solving in their curriculum.... Narrow tests...can become the de facto curriculum" (Resnick & Zurawsky, 2006).

Instructor comments about preparing their students for upcoming common summative assessments included the findings below. The numbers in parenthesis indicate how many different instructors out of 12 participants (10 primaries plus two pilots) addressed the topic/subtheme:

- “Teaching to the test” vs. “Teaching the test” (10)
- Practice sessions for students (4)
- Use of examples, “what right looks like” (3)

Looking at the impact summative assessment has on student motivation for learning, Harlen and Deakin Crick (Harlen, 2005) found that when preparing students to pass high-stakes tests is the focus, teachers resort to a transmission style of teaching (Black & Wiliam, 1998a; Linn 2000; Stiggins, 1999), and that students can be trained to pass any type of test, even ones designed to assess higher levels of thinking (Harlen, 2005).

“Teaching to the test” vs. “Teaching the test” (Learning Objectives vs. test questions)

“Teaching to the test” is a term usually used to describe instructors or curriculum that is focused on preparing students for a specific test, most often a standardized test. Here it is used to differentiate between teaching the specific questions on an assessment (teaching the test) vs. focusing on the learning objectives (teaching to the test). Popham (2001) used the terms “curriculum teaching” and item teaching.” He felt that because teaching either to test items or to

clones of those items eviscerates test validity... item-teaching is reprehensible and should be stopped." Ten instructors addressed how they taught lessons knowing what questions were on the common summative assessment that their students would take.

Hank described how he consciously tried to not item teach, or teach the test. "Probably not the best answer, but, the challenge, what I normally do, is I will look to see what the assessment is. The assessments here at CGSC are standardized and when you go through your classes, you want to do some kind of formative assessment as you go through, to prepare them for the summative assessment at the end. I do find it challenging that the learning objectives are so broad, that you can't look at the learning objectives, and then believe that you can follow learning objectives and prepare them necessarily for the summative assessment at the end. I think that's just something that we probably need to refine a little bit. But in general, I will look at the summative assessment as this is the end state we want them to get to. And then I will make sure that whatever the practical exercises that we assign in the class has the flavor to prepare them for the summative assessment at the end."

Hank continued, "There's not a yes and no answer to this and it can be bad, if what you do is... cover what's on the summative assessment, because ultimately, we want to prepare them for their next unit and being successful in the next unit, not necessarily the summative assessment at the end of the block. There are some that... [are] never going to touch the information that we're covering in class again. And so they're sponges as to what you give them, and they're the ones that will probably spit back exactly what you give them on the test. So, there's danger in that if you teach for the test, those guys, that's all they'll probably take out and learn. For those who have the background, who could probably do the summative assessment right up front and do OK, well, I would prefer not to teach to the test. So there are dangers in it, and I think the dangers are for those who don't have the background in it. So sometimes I am aware of that, I will have to craft what I'm doing on the board so it's not a replication of the summative assessment. But I do think there's a danger in it, because if we want the students to learn, if you're going to give them the answer on the board, well if they took good notes, all they're really doing in the summative assessment is just regurgitating their notes, and you're not assessing if they learned anything, you're just assessing if they took notes. So, that's the danger. If you teach too much to the summative assessment, then that's what you will get as an answer, and you're really not testing if they've really learned anything."

Frank was clear, he teaches to the curriculum, not to the test. “I don’t teach the exam, per se, and I think generally in DTAC, our assessments are linked with the instruction, in terms that they [assessments] come back and hit the key points and the learning objectives in the lesson. What I do is, particularly when I go through the lessons themselves that build up to the assessment, I make sure I hit the key points and make sure that we have an understanding of them. Particularly as we go through the lesson and then as we do the practical exercise afterwards. And usually the practical exercise is designed to address those same key points that they’re going to see on the exam.”

Frank continued, “I try not to directly teach the exam, as opposed to, saying ‘OK, these are the things that are going to be on it,’ and [just covering] those. But I do try to, through the questions and the discussions, try to make sure they understand some of the key points on the exam. Because for some of them... you have to understand the concepts involved in order to take the exam. Because [for the student] a lot of it is, I understand this concept, but in the context of this situation, this is how I would apply it. Particularly in terms of the understanding and the visualization pieces.”

Fred focused on teaching concepts, not items. “Well in a perfect world you wouldn’t have to preview what was coming up. The lesson material ought to sort of stand on its own in terms of, you taught the lesson, you have them [practice] it in class, they bring forward all of that to the test. There are certain lessons that aren’t going to be tested at all, by [course] design. So there is a little bit of foreshadowing over, which blocks [of instruction] have been structured as more important than others, recognizing that students can’t focus on everything equally well. For example the DCSA class, which doesn’t show up in any of the testable material, or the training management stuff in C500, which doesn’t show up in any of the testable material. In terms of weighting people’s preparation, I do foreshadow on that a little bit.”

Fred continued, “What I think I do is, if I know it’s going to be something on the test, and at the end of a lesson I feel there are students that aren’t getting it, then I will carve out time for those things that I know are on it, to make sure that the stuff that I know is testable is adequately covered, and that we’re not going to have people falling behind because they didn’t get that. Well, I mean certainly if you’re giving examples of the questions so [students] know if you’re essentially giving them a parallel question beforehand, then I think that doesn’t meet the point. But in class [if you see] your students are struggling with a concept that you know is coming up

on the test, then I think we owe it to them to make sure that the concept is clear to them before they get there.”

Caleb was clear, he does not teach the test, he teaches the curriculum. “I try to provide a context. The assessment, if written correctly, is going to have a direct link to the learning objectives. Sometimes that link is not there. So I’m going to focus on, here’s the material, here’s the subject, here’s the context, to better prepare them for whatever gate, or wherever these assessments are on the timeline. But, I do not teach the test. I will review and then I will try to by rhetorical questions or dialogue, get them to put it in a broader context from an organizational perspective.”

Caleb expanded greatly on the idea of a correctly written assessment, “The very first thing I would do, and it’s going to take a concerted effort, is to clearly define, we’ve got the learning objectives, the broad learning objectives that are owned by the college, or the school, or whatever, which has certain, has standards and, within it, which we are responsible for developing them, enable and learning objectives and the standards associated with those. The very first thing I would do, is make sure that those enable and learning objectives are clear in the skills and knowledge we want to assess, whether it’s formative or summative. That’s number one. So, we’ve got to clean up some ELOs. Number two, once I’ve done that, don’t develop the lesson plan, develop the assessment plan. What do I want to assess, in what perspective, and how does it relate to the ELO? For example, I’ll use the communication, for example. We have the broad, and I can’t remember what number it is, we have the broad, terminal learning objective of effective communication, or effectively communicate, which has two sub-sets. One is speaking, and one is writing. Well, if we’re, in our particular curriculum, and DLRO has the same thing, and uh, and some of the other departments have the same thing, if we want them to write a specific thing, for example, “write a commander’s intent”, ought to be a specific ELO, we all know that that’s an output and an input to the operations process, which is one of our learning objectives. However, if you want them to demonstrate that skill and build it over time, the assessment, the learning, the enable and learning objective has to be very specific in that, and, the instructions on the assessment have to be the same. It’s that linkage between our desired end state of knowledge or skills that we particularly want as a department, and it, and it changes at different points, you know, ILE common core has one purpose, and AOC has a different

purpose. I would start there. And then I would not write a single lesson plan until that was nested.”

Caleb concluded, “That’s why in my opinion, the challenge we’re having with some of our assessments is, we’ve got some broad learning objectives, which is OK, but when it comes to the assessment plan, you’ve got to have the learning objectives throughout and the lesson plan adjusting the teaching plan, but that assessment plan has got to be clear for the instructor in terms of the standards. And, for example, if you look at speaking, writing, is it organized, is it correct, all those broad things, but that doesn’t necessarily fit all the things we’re trying to get to in terms of effective communication so that assessment plan has got to be very finite. Some rubrics are very good. For example our COA Sketch and Statement rubric, it has some broad stuff, but it has kind of a scale, because a lot of students want to know why they lost points, and so, there’s a breakdown, based on doctrine specific things for the sketch, and there’s specific things for the statement and their correlation that have specific values on it, so that’s an example of a fairly decent rubric.”

Archie started by describing how he focused on teaching concepts. “I do a couple of things; I have a couple of techniques [I use] at the start of a block of lessons where I know I’m going to have a summative assessment, or several potential summative assessments throughout the block of lessons. Up front, reviewing the entire block, I discuss key decision points, key events. ‘This is a summative assessment, and this is a key event, and it’s going to be given out to you on “X” day, and it’ll be required to be returned.’ So I kind of try to set the stage up front. Then, normally, I try to do at least one, but normally I do two, about 20, sometimes it bleeds over to 30 minute, events in my classroom where I kind of do a, ‘All right, let’s just stop, and look where we’re at,’ in terms of the lessons that we have covered. And, you know, kind of the old trick, I’ll stomp my foot a little bit, [meaning] ‘hey, this is important to remember, this concept is critical that you understand.’ And maybe ask them, ‘do you the staff group understand why I’m saying this is a critical concept?’ If I can get them to be included on that dialogue, and how whatever concept I’m pointing out ties back to other concepts, or ties forward to upcoming events, we’ll have a dialogue about that. Those are my two techniques to insure that they understand we’ve got upcoming summative assessments and these are the key things you need to clearly understand and be able to articulate.”

Archie continued, explaining, “I have mixed feelings [about teaching to the test]. The idealist educator in me says, yes, teaching to the test is bad. I want them to grasp the concepts and then through classroom dialogue and understanding be able to develop their own mental models of how those concepts fit together and be able to articulate that clearly to everyone else. However, I understand, based on time, time being the critical factor here, the time we have in the classroom, and experience, the experiences, branch experiences of each individual officer, that... we may not achieve that ultimate end state, where you teach everything without teaching to the test. There may be some critical things you have to stop and say, ‘all right, you’re going to see a question similar to this on the exam,’ I don’t give out ‘the’ questions to the exam, but again it’s a little bit of nuance in gauging where your students are in terms of experience and gauging how much time you have available, and whether or not they’re really grasping the concepts that you want to impart to them.”

Allen said he does not “teach to the test.” “Provided the test is set up to where it’s actually testing [students] on what the leadership in the department or the school has said, here’s the key points that everybody should be walking out of here with, so okay, if it’s set up that way and it’s supposed to measure their knowledge of the key points that they’re supposed to be understanding when they come out of a block, then ok, we got to make sure we [cover] that. To me, it’s like driving down the highway, it’s like yet another road sign to the instructor to say, okay, make sure you do these things.”

Referring back to an earlier question, Allen explained his views on effective assessments. “OK, what are the key things that we want the students to walk out of each block being able to know and do? Have we built the assessments to assess their ability to do those things? I don’t know that in all the cases, with all of our assessments that we really do that. Because that... kind of makes everything go in line right there and it helps to re-emphasize with me as an instructor, OK, here are the key points you’ve got to make sure you cover with them, and then we’re going to grade them on that and see how they do overall. So, I think that would be a thing to improve on.”

Alex “I make sure that, for any assessment, I talk to them, I take a little bit a class time and say, ‘this is our end state, this is where we’re going, so understand where that is.’ And that’s not saying that I’m giving the yellow sheet [assessment document] ahead of any faculty or anything, but, here are the key things that you’ve got to know as a learning outcome that we’re

supposed to have and we're going to talk through these things. So I make sure that they understand it ahead of time, and then, while I will not give the specific, the questions and things that come out of those assessments, I do make sure that I look and see, OK, if we're going to assess A, B, C, and D in our teaching curriculum, have we touched on A, B, C, and D, and sometimes you find that there's gaps and seams, and I think that's my job as an instructor, making sure that I've at least touched on it. Because if that's what the GPA is going to be based on, then I at least owe them that."

Alex continued discussing teaching what's on the test. "I'm sure that I have. I don't know that I recall a specific, but I don't see how you could not. Just from a standpoint that every one of us has our own things that we know we want to hit, and I think one of the things that you have to do as an instructor, is, you have to assess your own staff group, and where their strengths and weaknesses may be, may not necessarily line up with what the assessments going to be. So, you know that it's something, it's an area of weakness for the entire 16 [students], you want to spend some time on it, but it has nothing to do with an assessment. So that's important, but I also, again, to be fair, I've got to make sure that I at least come back in. So I don't know that it's teaching something just because it's on the assessment is necessarily a bad thing. Sometimes it's just, I've assessed there's other areas where I want to expand some depth on, but I can't just forget that, so."

Addressing "Would you spend more time on that than you normally would have because it's on one of the assessments?" Alex answered, "I probably spend a decent amount of time on it, just because I find that most of your students don't understand it, and don't know it or have heard of it before. So I think it's useful as long as you're giving it the context of, a defeat or stability mechanism is tied to your operational approach. It's really more on the design methodology and conceptually how do I want to solve the problem. The other thing that I spend some time with that is, in particular, defeat mechanisms we can have some very good discussion on the difference between accuracy and precision of the universal and actual standards because defeat mechanism has words that are also compatible with tactical task, but they don't mean the same thing, so the context, so there's, I think there's some goodness in that."

Dexter described how he taught to the curriculum. "I don't do it the way that I have a perception that some people do. I don't read the exam deliberately, or assessment deliberately and know exactly what the questions are and then tailor my instruction to ensure that I have

personally provided them the answers. Meaning that I don't use the questions on the exams to tailor my instruction. Whereas, I know some people do. I focus on the learning objectives that are part of the block, the lessons themselves. Preceding the student meeting, I will look at the assessments and make a personal judgment as to whether or not I feel that either through their individual, outside of class requirements, meaning in their study, their reading of assigned, assigned requirements, since they're all open book, there's the opportunity for the individual student to pursue the answers outside of from pure memory. The [the assessments] are not knowledge based, [students] have the opportunity to do their own individual research and study. So, that's how I do it... I don't foot stomp the answers in the classroom."

Dexter continued, "My philosophy is that there's the aggregate effect of their personal individual responsibility regarding the course material. It's not a secret what the pre-class reading and study requirements are. Those are well written and very available to the individual officer. I work on the assumption... that the students have done at least a level of preparation to come to the classroom in order to contribute and participate in the discussions. The one thing about our summative assessments in some of the examples that you've got highlighted here (points to assessment classification form), we generally give them at the beginning of the block, so that as they negotiate or navigate the lesson activities or the lesson periods, if they're a smart or thinking individual, they're familiar with the questions. When we're in a particular lesson period or a block that is either, overtly, or very clearly related to the, particular question, then if they should posture themselves and seek greater understanding, greater knowledge, greater ability to provide a thoughtful, complete, and professional answer relevant to the way the question is provided to them."

Jack focused his students on the course learning objectives. "The framework in my estimate, what the students have to look at to prepare themselves, [are] the learning objectives, for the course and for each module. [They] are fairly straight forward in what [students] should be able to carry out of the class. I will frequently give them quizzes, perhaps, or if you will, an oral discussion during class to see where they're at against some of those objectives."

Arlen, like the other participants, focused on teaching to the curriculum. "I think, again you want to reference the material, but I don't think giving them the answers as part of the instruction should be there. I think it should just be included in the facilitation and discussion, but it's up to the individual student to learn, to check on their learning and kind of evaluate how

much they read, or how well they understand the issue, and giving them the answers is a fail as far as I'm concerned, on the instructor level. I try ...not to spoon feed them, and say, 'these questions are going to be on the test,' but I more or less try and make sure I at least go over some of the topics that'll be addressed on the test so they have a familiarity with it. I'm not going to give them the answer, but I'll kind of go over it and we'll talk about the topic in general terms, and then it'll be up to the student, based on what they're reading is and understanding, to do it on the exam. But, I don't think I change the teaching style, I just make sure I at least try and include some of that discussion on topics that are on the exam in the classroom discussion."

Teaching to the curriculum instead of teaching the test was typical of DTAC instructors, an indication that instructors saw their role beyond just preparing students for upcoming assessments. Several instructors help practice sessions outside of class hours for students needing additional assistance.

Practice sessions for students

Four instructors discussed practice sessions focused on topics that will be on an upcoming assessment held prior to students taking the common summative assessment.

Allen would identify students he thought needed help and offer to work with them outside of class. "If somebody was really struggling, and I was picking up on it in class, I would pull them aside and give them the option of working offline with me. That's probably been a big lesson that I've learned from this first year of actual teaching, was, from talking to everybody else who's been teaching for several years, is OK, here's different tools that you could use as an instructor for those kinds of things, to give them some additional things that they can work on their own, and then bring them back in and get some more feedback to them."

Arlen offered assistance after class to all students, but usually the attendees were "...some of the international students, and the interagency students, usually not the military. I do offer it to anybody, but usually the ones that take advantage of it are the non-military students, and I'll give them some ...additional help, and we'll work some problems, associated on the steps, just so I know they have the understanding. It's open to anybody, but the only ones that have really taken advantage of it are maybe the non-army guys. Air Force, Navy, or interagency students have taken me up on that option, but it's open to anybody."

Alex met one-on-one with students. “I’ve had individual students say that they’re having trouble understanding something ...sit down one-on-one, but I’ve never done any type of extra practice sessions.”

Jack scheduled extra practice sessions for groups or individuals that wanted to practice on something, “Usually, course of action development, commander’s intent. I’ll offer it, after class, for any students who are interested.”

Four instructors described meeting with students after scheduled class hours to practice skills from the DTAC curriculum. Several instructors used examples of “what right looks like” to help students understand requirements.

Use of examples – ‘what right looks like’

Three instructors discussed using examples of “what right looks like,” during instruction prior to students taking a common summative assessment.

Fred focused on examples that doctrinally based, preferably from a doctrinal source. “I definitely point out on the slide material when there’s some examples that are better than others, so we do work in class to critique how is it that we could make that example better. We show that as an example of a commander’s intent for example. Here’s all the options, or how we might tweak that one to make it better... I don’t tend to provide a lot of outside examples in terms of this is what right looks like, outside the doctrine, because I like to keep that as the basis at least for C500. O300 is a little bit different because we’re getting into some more advanced stuff, so I might pull out figures or products from another [Army] unit that I think is particularly well done. But early on when we’re doing the doctrine stuff, I think anything that takes away from the actual [doctrinal reference] manual is shifting from a known point too far.”

Arten used examples that did not directly match exam questions. “For example, specifically in the course of action statement and sketches, we would definitely go over a couple examples that are similar in the offense and defense to give them a frame of reference. I think, especially in C500s challenge, because this is the first time some of them may have seen a sketch or a commander’s intent where, you know, they’re just not familiar with it, so we do give them some examples, but I’ve tried to ensure the examples do not match directly the exam questions, but yeah, there are some examples that facilitate discussion on it. I think that’s helpful to the process.”

Hank focused on teaching concepts and problem solving by using examples of work. “A lot of the students want to know what does right look like so, I’m cautious about giving them what I believe right totally looks like before they do the summative assessment, because normally the summative assessment is where we want them to take some of the concepts that we’ve discussed and think about it, analyze it, and put it together, because that’s field grade [officer] development, it’s being able to analyze things, and come up, formulate a solution to something. That’s problem solving. If you give them here’s what right looks like, then generally there’s a danger... if you give them that, then that’s exactly what they will provide you, as their analysis, is what you gave them. And so, have I done it in the past? If I think the concepts are challenging, I’ll try to use a different scenario. So I’ll give them, here’s what I believe right looks like, fully knowing that there is a danger in that. And so, I’ll have to look at the summative assessment, and make sure that I completely change the scenario, or I’ll completely take a different angle at something, so it’s not a similar scenario or something like that.”

While preparing students for upcoming summative assessments, “curriculum teaching” was typical among DTAC instructors. Other preparation practices included holding practice sessions and the use of examples of “what right looks like.” The next section completes the findings of similar practices used by DTAC instructors as a result of using common summative assessments by exploring the relationship of the grading process.

Theme Four: Use of the Grading Process

“Grading infuses everything that happens in the classroom”

(Walvoord and Anderson, 2010b, p.1).

Instructor comments about the grading process and common summative assessments included the four findings below. The numbers in parenthesis indicate how many different instructors out of 12 participants (10 primaries plus two pilots) addressed the topic/subtheme:

- Student motivation – prioritization (11)
- Stratification of students (10)
- Calibration amongst teaching teams (5)

Ultimately, the responsibility for evaluating student officers falls on the individual instructor in the classroom who is charged with the dual, sometimes competing tasks of evaluating and developing student learning (Ewell, 2002). Suskie (2009) explained that the grading process is part of assessment; “Assessment, especially the grading process, motivates students to do their best” (p.59). Reeves (2011) characterized grading as a form of feedback, adding that it is also a very powerful instructional technique when it comes to influencing student achievement.

Student motivation – grades/prioritization

Whereas grades are often used to control student behavior (Rowntree, 1987), a formative approach can counteract student’s obsession with grades and redirect interest towards learning and, if students see summative assessment as formative, they will accept it more (Black, 2004; Black et al., 2004). Eleven instructors described how they felt about the relationship between grades and student motivation. Most preferred to describe student activities in terms of “prioritization” over motivation. Walvoord and Anderson (2010) noted that grading affects how students study, what they focus on, how much time they spend, and how involved they become in a course (p. 2).

Frank felt that summative assessments motivated students more than formative assessments because of grades. “I think that summative assessment probably motivates [students] a lot more across the board than the formative assessment. Particularly when you get to some of the marginal students. I think the students that are in the top 50%, they’ll approach them equally, but I think I’ve seen some of the ones in the bottom 50%, will put maybe less effort into something they know is a formative assessment as opposed to a summative assessment. I don’t know, it’s a matter of time and effort and payoff that comes into the decision-making. But, I do see a side difference.”

Later during the interview addressing prioritization, Frank added, “...probably the bottom third probably put less effort into it. You know, the top half, top third, are going to approach all these similar, but the other ones you could see, in some cases, they took some risk.

This comment by Frank, used previously in another section, is included here because it addresses student prioritization. “...I realize the problem that we have in timing, particularly that

our curriculum has to tie into what's going on at the TASS (Total Army School System) sites and the satellite campuses, and that sometimes drives our timing and affects what we have to do. Also, we have the other departments, you know, what's going on student-wise with all the different assessments. So, we have to balance it out so you're not hitting them with everything at the same time, so that they have some time to put some quality effort and reflection into, to their assessments and into their work."

Fred's initial comments addressed students just getting through course requirements, "In general, there are some students I do think pay a lot of attention to the, to the feedback, I'd say probably, I'd say like 25% of them will come back and ask you questions about the things you wrote. The other ones, as long as it's not a failure, stick it in their bag and leave. You know, as long as it got them through."

Frank added at a later point in the interview, "They all want to do well, and so they get frustrated when you have them do something that they know they're not prepared to do well on [including] ...our take-home [assignments], where we basically give them an open-ended period of time to do it. We give them three or four or five days to do something, I think they all wrestle with how much time I need to spend on this to do well, relative to my peers. I think that gives them a fair amount of stress. I personally think if we just told them to be in a classroom for three hours, here's your block of time and there's a start date and an end date that would relieve some stress... It would raise the stress on some of the under-performers but, I think the vast majority of them I think would be relieved by that because they wouldn't [feel] they just needed to keep on working and working and working. Because I think really time management is the motivator for most of these students. How much time do I have to put into any of this, versus all the other things that I'm doing? I think that's really the limiter on their extra achievement."

Frank followed up later, "I think again that depends on the student. I think there's probably 60% of them that essentially view the course as pass/fail. And the grade is not a driver. I think there's another chunk of them with the motivation to be in the top 20%, to be able to clearly demonstrate that they're ahead of their peers, matters to them."

Archie addressed grades as a motivator. "I think just giving them the straight summative assessment [without a formative one first] at the end, kind of just dampens their motivation to do well. When you think about the students partially, most of these students are high achieving, relatively intense individuals, expect a lot of themselves, and most in the past have done well

educationally, for the most part. And so, when they don't do well here, it's kind of a shock to them, a little bit. And I think that adds an additional dampening effect on their motivation too, to perform."

Archie continued his discussion on grades as a motivator, "Not here, no. I think grading played a little bit of a factor in the summative assessment exam, the C500 exam, because while some students felt they grasped the concepts pretty well, when it came to demonstrating their ability to grasp the concept they didn't do as well as they thought they understood the concepts. So it became a little bit of a shock to them I think."

Caleb began by discussing addressing the impact of grades on students. "That's a tough one, but I'll give you a perspective. I think for the most part, students, when they receive that feedback, it either confirms for them that they're learning it or they understood what was being taught, and most students, the majority of students from my observations, will take that to heart and make adjustments accordingly. Or if they have a questions they'll come see you and say, 'hey I don't really understand this, what did you mean by that?' and so it increases communication on an individual level between the instructor and the students. But it also, potentially, depending on the maturity of the officer, and I saw this happen twice initially during the early part of the curriculum with a couple of my officers, it also gets to some other things beyond the assessment in terms of, some individuals equate their value to whether they got an A or a B. So that also enables communication. But then that's when you have to reinforce, you're not dumb, you may have thought you were Napoleon, it's just that maybe, you know. I've had one officer come to me who already had a master's, very competent officer, but, in terms of tactics, and the stuff that we teach in ILE common core, was a solid officer but wasn't doing "A" work, in terms of the art and the science. This officer questioned their value in terms of; well I've never gotten a "B" on anything. And so, although it enables feedback, there's that perception of, in some officers, that their value is equated to the grade, it isn't about the learning. So, I mean I think from an assessment perspective, that's good that we as instructors are aware of how they learn and how they value those grades. For the students, that's one way in terms it enables communication, but the other thing too is, in terms of their professional development, the majority of students, based on my experience, will take that as kind of a benchmark, and say hey, I'm going to try to do better next time, or maybe I need to connect this better. I've seen over the course of the 40 weeks they're here, the ones that apply themselves either self-study or they will

continue to get better in that particular area. They may still have some other weak areas, so most of the students will use it as a benchmark to, to keep them on glide path.”

Caleb continued, “I don’t think its motivation. I think its prioritization. I mean, common sense would dictate and I can’t say this for sure because we only have really one formative assessment this year, but they didn’t stack arms [quit], or it wasn’t like I’m going to put that lower or hey I don’t have time to do that, I won’t do [that] in terms of it’s the right thing to do, but I think, it’s not about motivation, it’s about prioritization. For example, a student, regardless of whether it’s summative or formative, if they’re in a master’s program, where there’s an MMAS or an outside [requirement], they’re going to prioritize their work. And I have seen that my normal “A” student, who is doing great, because they’re doing other things in terms of outside work or whatever, will drop maybe half a letter grade because they, and they tell me straight up, Sir, not my best work, I had to focus on this, but it wasn’t a motivation thing, it was a prioritization thing.”

Allen tied in the timeliness of feedback with prioritization. “I think it depends first of all on the time of the year. Early C130 can be kind of a shock to their system that, ok, wow, I didn’t know everything that I thought I knew. As the year starts to drag on, especially if they’re repeatedly doing poorly on things, or it’s a time-frame where they’ve got multiple assessments from multiple departments that are due all in the same week time-frame, or something like that. Then, overall, the motivation can get kind of low. I think an important thing is getting them the feedback, back to them as quickly as possible, because they’re on such a quick pace academically that you go beyond a couple weeks it’s, OK great, yeah, whatever, I’m onto the next three sets of assessments that I’m doing... I think with a lot of them, it probably takes more effort to try to remind them of, OK, here’s where you had issues with the previous assessment, here was the feedback that you received on it, we’ve got another block of stuff that’s coming up, take that all into account as you go through and do your assessment for your next block here.”

Allen later commented on motivation and grades. “I think when they hear formative; they’re a little bit less motivated than when they get told it’s a summative one, because to them formative means, OK, I’m really not being graded on it. The 7th Azeri COA Sketch and Statement, I think that they were finding so many extra things it was leading to some frustration on their part. The original C533 COA Sketch and Statement, it was a good thing for them to get, but it was at a point where they didn’t know what they didn’t know, so to speak.”

Allen felt that some students who wanted to just pass thought his grading was too hard. “When they got their feedback it was like, wow, I thought I did good enough to get the Leavenworth “B,” and that isn’t really what I got. So, I think, sometimes some of them will get that attitude of, wow, you are really being nit-picky on these kinds of things, where, no, really we’re not because all I’m doing is taking the [doctrine reference], and basically going through the list of things that are supposed to be on here.”

Alex explained student prioritization vs. motivation. “I would say that to the majority of students I don’t think it has a motivational impact. I think my experience is most of the students come here wanting to do well, and they know they’re going to be assessed and graded, and that’s just accepted. That’s what goes with it. Where the motivation becomes a challenge for them is when five different departments have five different gradable assessments due virtually all on top of one another. If we’re doing a summative assessment on Friday, you know, we’re doing it tomorrow, but during this week they’ve had a leadership paper due, a history paper due, and ethics paper due, well, let’s all be honest, they’ve been closing out those, and our summative exam has been the 300 meter [close target], then the 250, then the 200, and so they probably haven’t reviewed too much, except for maybe just the night before, and then they don’t do well. When I say they don’t do well, maybe they get an 85, which is a Leavenworth “B” which we say is the norm, but if it’s a well-motivated student that is not happy that they get anything below an “A” minus, that’s a motivational issue, but it’s also, they’re trying to prioritize so many things, so I haven’t really seen it as a motivational issue if it’s out there by itself and they can focus on it. Now, there are certainly others that, I think if you’re motivated because you’re scared about it, you’re worried about it, there are probably some larger issues with you as a student anyway, and you’re probably a weaker performer. I mean, you’ve got to assess that, that doesn’t mean you’re going to change how you grade it at all, but. I just haven’t seen it as a negative or positive motivational.”

Alex described his students who wanted to take an assessment ungraded and receive feedback before taking it for a grade. “Very positive feedback from the students, they appreciated that. In fact, the AAR comments that I got back at the end of O300 was, they would have liked to have seen the Statement and Sketch, not the Commander’s Intent and Planning Guidance. They would have like to have seen just the Statement and Sketch come back, not necessarily exactly the same one obviously, but same task, same requirement, but be graded, so

that, they have a free shot, they've got instructor feedback, they've gone through some additional class work, now let them show that they can perform the task to standard or not. Same requirements, just give it a different scenario and now this time it's for a grade."

Alex later addressed outside requirements and prioritization. "To a degree, yes, but, I never try to make it a point in the class of where, how you graded vs. someone else. I always try to make it a discussion with them of how they graded out with where they think they should have graded out at. We need good, healthy competition, but I've never made the top 20% a thing that I push to the students. I think that if that's something that they're seeking, let them identify that it's important to them, and then let's have the discussion, you've told us in your IDP, you've told your coach that you really want to compete for the white brief case, but your class work is not up to that, so let's talk a little bit about what is it you don't understand? Are you also doing another master's, outside of the college? You've got stuff going on at home and if you don't understand then we've got something to work on. The rest of them you control. So if you're telling me that it's important for you to make the top 20%, then you need to adjust some of the things that you can control."

Alfred (Pilot) felt that competition amongst students was a motivating factor, "Especially among the meat eaters (the combat arms officers) in the class. The folks that come in outside of the combat arms branches, especially during the tactics block, I don't know that they feel pressure; they may feel pressure to provide their level of expertise to the discussion. But, you know I think there's probably a, a good, healthy competition. I don't know that pressure is a good word, but maybe a healthy competition to do the best."

Dexter equated motivation with student level of effort. "Well it's hard to judge, I mean, if you're going to be evaluated, regardless of whether or not it's a formal grade that's entered into your cumulative or total grade point average or it's [non-graded] to me the expectation is that there, again it goes back to the idea of level of effort. If you have a requirement, the expectation is that you do your best. I tell my students there's a difference between perfect, good, and good enough. We do speak about, value what we say, and that is balance. But, if I were an individual officer and I knew that the grade was not going to have a significant impact to me, well I'm going to do it to challenge myself, to show that I know what I'm talking about, then I'll put an appropriate level of effort to make sure that I communicate or show that. But, I'm not going to worry necessarily about whether or not I got a 95, or a 98, or a 100 on that particular

requirement so, my level of effort will be... there are a lot of variables that will go into my level of effort. Summative assessments, I think that we kind of got the guidance to power-down the average grades, meaning go from average being low to mid-90's, to a mid to upper 80's as the mean of grades. Personal choice by the individual officers is the bottom line.”

Later, discussing the effect of top 20% designation on student motivation Dexter added, “Well, I don’t know anymore. I think that it’s different now than it was two years ago before we re-introduced the top 20% award meant for [those who] exceeded course standards. There are the natural top performers who will do well regardless. There are those who will probably ratchet up their level of effort based on the school’s decision to re-introduce the top 20%. And then there are those who will just ignore it, but I guess the answer to the question is, what is the impact of summative assessments?”

Jack did not think grades were a big motivator. “I don’t know. I haven’t seen, I haven’t observed a lot of students who are focused on letter grades. Most of them I observe are focused on learning as much as they can. Some are not good test takers, but are very knowledgeable.”

Arlen thought some students were motivated by grades and saw assessments as a way to focus student effort. “I think definitely, some students are motivated and want to get the “A”, we have a lot of type A personalities that think the assessments are important and it focuses their study efforts. I’d like to say that’s the majority of the students, whether it’s 60% or 80% of the motivation, of the student population of the class, the higher the better. But definitely a motivator to make them understand and again it’s a check on their learning to see how well they did.”

Arlen added to his thoughts on graded vs. non-graded assessments. “Oh, huge difference, I mean if it’s a summative assessment, they’ll put more effort into it. The formative assessments, I encourage them to put the effort into it, but there were a couple students that did not [they] did the minimum to get by. I’m not getting a grade on this, just getting feedback. I think that’s where you need to mentor and coach the students and say, I know this is important to do, important for you to understand, but the summative evaluation, assessments, for the majority of people, the students put more effort into those than the formative assessments.”

Hank “Well, some of them are pretty good at making the adjustment. I think that, for the most part, DTAC tends to be focused more on the, the MFE officers [combat arms] obviously, specifically the maneuver guys. So, those guys tend to, at least in my mind, they’re the ones that are going to ask me a lot of questions when I write a lot of comments on their stuff, because they

know they're going to have to do it again in their units. So I think they tend to take it a little bit more to heart. I do have some that are just good analyzers... lawyers tend to not have a lot of corrections on their stuff, and so, I would say in the last three years I've had a lot of discussions with lawyers, and they tend to make adjustments based upon my feedback to them for the next time, so they're good analyzers. Lawyers tend to be really good at analyzing things, and they'll express themselves pretty darn well, and then the MFE officers, they're motivated more because they have to do it again. Somewhere in the middle are those that, I would say that, they'll take my feedback, and they may refer to it for the next time they have to do something similar, but I'm not sure I see huge steps forward because sometimes I think it's more of a check the block. It's not their bread and butter, it's not what they're going to do in the future, and so they'll make some adjustments, but it's spotty at best. There are groups that I know that will take it more seriously than others."

Hank continued, "There are motivations when you give grades. There is no doubt about it. If you are pressed for time, and it's not graded, well, the level of effort you're probably going to put into it is relatively low. You have to do something to kind of motivate them if it's, hey, when you come in, we're going to present your COA statement and sketch, well, that's a motivation. It may not be graded, but at least that's a motivation to give it a good shot."

Hank later discussed prioritization, "We load them down with a lot of summative assessments, and a lot of times, because there's so many different departments, they tend to be back to back or right on top of each other and so what I believe the impact is, at that point they don't have a time to sit and think through, and always come up with their best answer. A lot of time they're giving an answer based upon time management; well, that's good enough, let's move to the next, because they're being rushed because they have so many topics, because there are subjects that they have to read for and prepare for, and because they have so many other assessments going on. I do believe that's an impact on summative assessment; the amount of stuff that we pile on them. I think there is a positive aspect to this summative assessment, because they kind of want to know [where they stand]. Most of the time summative assessments are individual assessments and this is one of those few times that you can talk directly to them, and say, hey you got it, you are right where you should be, or hey, you're not quite there and here's some of the challenge that you have. Because a lot of stuff we do in class [is in] groups, most of the grading we do tends to be group, and so I think these summative, individual

summative assessments are a good, positive thing for their feedback. But, like I said, there are some other things that get in the way and I do think that we tend to add too many things on to where they say OK, that's good enough, 80% is good enough, I just got to move on. I have students that I would call them my, efficiency type students; B and go, because they've got to get on to the next subjective."

Eleven instructors described how they felt about the relationship between grades and student motivation. Most preferred to describe student activities in terms of "prioritization" over motivation. The stratification of students during the grading process was typical of DTAC instructors.

Stratification of students – diverse population

Bloxham and Boyd (2007) contended that differences in instructor approaches to grading assessments, categorized as norm-referenced or criterion-referenced, often result in unintended outcomes. Norm-referenced assessment is designed to distribute student performance over a range, for the purpose of discriminating between those who do well and those who do less well (Bowden & Martin, 1998). An example of this would be if 30% receive an "A," 60% a "B" and the remaining 10% a "C." Criterion-referenced assessment judges students against a set of criteria that is linked to desired learning outcomes. Bloxham and Boyd (2007, p. 82) maintained that "whereas with criterion-referenced assessment all students have an opportunity to do equally well, a norm-referenced approach will almost always create a distribution of grades." Ten instructors stratified their students in some manner during the grading process.

Dexter's explanation of the stratification of students during the grading process mirrors what most instructors felt. "I've had these conversations that it should be graded, the awardment (sic) of a grade should be against the standards expected for the performance on that particular evaluation, assessment, or individual particular question. Either knowingly or unknowingly, there's an introduced bias and/or, variable of, comparison of individual student to individual student, and we operate within the, the confines of our 16 person staff group, and that's part of the struggles of the school at this point. My individual instructor or faculty's judgment, comparing some of my students against each other, can be uniquely or distinctly different than another faculty member's individual judgment of their 16 students relevant to each other. So, as much as we would like to try and normalize the faculty's assessment and judgments of student

performance, I don't know that it's impossible, but at least currently based on the way the assessments are organized, I think it's hard to do."

Frank used the top third, middle third, and bottom third technique. "Usually what I try to do is, when I grade, I usually just go through all of them first time, and just sort of reading them, without putting any marks on them. Just sort of align them to see how they're doing. Make sure that, you know, they're within what I consider tolerable responses. And then based on that, I'll sort of break them into top third, bottom third, middle third, and then I'll start going through the grading process and actually start to assess them and once I'm done with that, then I go back and make sure that I've been consistent throughout the grading process. Because sometimes, you have to guard against the tendency to start out too hard or too soft, and the guys that went first either get it extra hard or extra easy. So I want to make sure I'm consistent particularly if there is a recurring problem or issue throughout the exam that I have been as fair and as balanced as possible."

Frank compared students against each other as well as the rubric, "...I sort of use a modified curve, I want to get a feel for how everybody did across the board. I use the rubric for the guideline, breaking down the points, and then read all of them together just so I get a feel for, okay, across the board this is how we did, and then I divide, like I said, you know top third, middle third, bottom third, and then using the rubric, [look at] ...the way that their peers have done, and incorporate all those... and then I say what I think is a good answer, or good response into the final grade."

Fred explained how he used norm-referencing when grading his students. "...because to some degree I think the summative assessment is as much a review of how well I've presented the material. So I definitely go back and look, across the sixteen students, are there any trends that I can see across all the answers that would indicate that I'd either done particularly well on a concept, or that they don't misunderstood something I've said, or weighted things wrong, or, whatever the perception is. So, I look to see if I need to go correct and re-teach anything. Then also I do that again a little bit for fairness, just because there is some fatigue when you grade some of this stuff, or the first couple you grade, you hold to one standard. Over time when you see other people's ideas and you see how student 5, 6, 7, & 8 & 9 all did much better than that, so then I go back, well did I apply the same standard across the board? That's a difficult thing to

do, but I do try to make sure that I don't get easier as I go, or harder. Usually what I find is that I get harder as I go."

Fred described bracketing student answers. "I definitely think that's part of the process, and I try to do, when I grade it, I try to pick the student who I think will do the best, and then the one who historically struggles, and then I go back and forth between what I think will be good and bad. Partly because it's less fatiguing to do that, but also so that I can try to bracket in where I think the left and right limits of the answers are going to be. That doesn't always work, because you always have a student that surprises you and either under or over performs."

Fred continued, "I do use the rubrics, because in general they help talk about which points, how much points or how much weighting should be given to each part of the problems that we give them. Within each chunk, for example, if writing the commander's intent is worth 20% of that particular assessment, then I will go in and use the other half of the standard, which is the A, B, C, D rating that's published at the beginning that basically says a B, 85, is about what you expect of an average student to produce. So not particularly acceptable, it gets at the answer but it doesn't show any real original thought. I will look at the answer and say, OK, is it about what I expected? And if it's about what I expected, then they're going to get in the B, B+ range. If they've miss-read the question, if you can tell they were cut and pasting from something they read in the book and were essentially rephrasing something else but hadn't really added any meat to it, then they might get lower based on that standard. And if they produce something in a new way, or they connect the material to something else or other big ideas, or other parts of the curriculum like leadership or history, or pull in an example from another class, then that will get them in the A range. Really, I mean you really have to knock one out of the park to get an A+. It's got to be pretty original, perfectly presented, and really demonstrate that you've making links to other ideas beyond what was said in class, but really sort of your own study into it."

Fred discussed his thoughts on in-class testing vs. take-home which was more common during the case study time period. "I would like to explore more of them in class [students taking summative assessments in class rather than as a take-home]. The other thing about in-class I think would be useful as an instructor, as you proctor, I think we could get a sense from body language and pace on who is struggling as they do it. About how much stress level is involved. And, who is not really applying themselves very well. You know the guy who comes into a three-hour test, knocks it out in 45 minutes and leaves, is either a genius or doesn't care. And,

right now, I don't know how much time students are going and preparing. So, if a student goes home and spends three hours on the test and does pretty well, that's probably a good thing. If another student goes in and spends nine hours on a test, and is performing at the same level as the three-hour guy? We've got to do something about the nine-hour guy, because it shouldn't take him that long. But, I mean, I've got no way of knowing how much effort is getting the same result when it's outside of the classroom. And, and I think that would be useful for us to know. I mean, because when 'we kick them out of here', they won't have forever to work on projects, and we really need to be able to know who's performing at that level or not."

Fred later discussed how he felt about the top 20% academic student designation. "I don't have a problem with that at all; I think it helps us differentiate those who are above average. And, I mean personally I think that should be one of the goals of this institution, is to be able to report out to the greater army, here are guys who, either naturally get it, because they're the kind of person who just get it, or these are people who work really hard and are focused enough to be above their peers. I think those skill sets ought to be reported out. And if we've got people who just don't keep up, that needs to get reported out too."

Archie described stratification of student performance after the grading process. "The second thing I do is, when I hand the assessments back out, again it depends on time because you may be intruding on another instructor's time as well, and I don't want to do that, but what I'll do is I'll clip the assessments, after I've graded them, you know, based on performance, and I'll pull out three or four students at a time, hand their exams back, and talk in general terms. Here's what I saw on the exam generally from the entire staff group. All of us did this fairly well, all of us answered this question fairly well, and all of us struggled with this concept. Overall, the four of you performed above average, or performed in an excellent manner and exhibited a clear understanding of the doctrinal concepts. I have no issues. If you have concerns, you can review your exam and come back to me individually, or send me a note. And then I'll go to the next group of four, which may be, four or five folks who've performed, you know, to the standard. They may have struggled in one or maybe two areas on the exam. I kind of give the same pitch to them, but I'll emphasize the point of, I want you to come back to me individually, not formally, but individually off-line, catch me for five or ten minutes outside of class and just re-explain what you think this concept is, and we'll have dialogue about it. And then the last group, you know, are those students who have struggled on the exam, for whatever reasons, and I'll pull

them all out and I'll say, you know I'll just be frank with them, you know, hey, you guys didn't do so well on the exam."

Caleb described the art needed to grade assessments using a rubric. "...first of all I look at the rubric... some rubrics are better than others. First what I'll do is I'll read them all first, whatever they are, and generally see if they've got the gist of the stuff. Then I'll go back and it really depends on what it is, I mean, you've got to use the rubric, but if for example, in ILE common core, a lot of the new doctrinal concepts, the doctrine changed so much, so I'm really focusing on, initially, do they have the science, in terms of the rubric. Whether it says they're going, do they really, you know, do they understand the doctrine, are they using it correctly, that kind of stuff, are they consistent. Then I'll apply the art piece, based on my experience and others, okay, now, they've got the science, now they've applied aspects of the art. And those officers that generally, within the rubric, that apply the art within the rubric generally will get the higher marks in terms of "A", or whatever. Now the challenge isn't the grading standards, the rubric, sometimes the rubric is so spongy, that you have to also use other criteria that may not necessarily be on the rubric. For example, our broad ELO is like speaking and writing effectively, so did they clearly articulate, in writing? So whether that's in the rubric or not, you're going to give them some feedback on the writing. Did I understand it? Is it coherent? Is it in the right format? Whatever the guidance was, either changed or otherwise, did they follow instructions? But more importantly, did they answer the question? So in terms of the grading, generally those officers that have answered the question, can clearly articulate their thoughts, and have gotten beyond the obvious answer, you know the critical thinking, the higher thinking, the connecting the dots, those generally will be in the higher end."

Alex described using a criterion-referenced approach. "I rely very heavily on CGSC Bulletin #903, which is academic performance graduation awards policies. "Each one of our class advance sheets assessments, particularly the advanced books will have, OK, what's an "A", and it's a lift out of this, so that's where I start with is, looking at those words of what's, what's the difference between satisfactory and very good? Between a "B" satisfactory, and a "B" plus of very good? Okay, meets the standards is a "B" plus, meets most of the standards is a "B," and so, everything I've ever graded I go back and I start with this; that's where I start my calibration at. If you read how it's written, we say every student here should be capable of a "B," then we're saying that, by the school's definition, all of our students can meet most of the standards. They

have to get a “B” plus to meet the standard. So, there’s a little bit of that. So I’ll start with that. Some assessments, from the curriculum design are very clear, you know, how much each part is worth, and I’ll use one, the Commander’s Intent in O320, and I can’t remember how it was set up this year. There are three parts to Commander’s Intent. If you tell me it’s worth 10% of the overall grade, out of 100 points it’s worth 10, I will sit down before I start grading, I will break it apart and say, OK, well then, the three parts are worth 3.33 points to me. And I will even look at, OK, an end-state - they have to address how the friendly force stands in relation to the enemy to train and to civil. So, that’s subdivided by three, and so, and I’m not normally an anal guy, but I am sort of anal at how I look at that, because, while I am subjectively grading how I think they did, 99% of everything we grade as a department is subjective. I know that I am not the expert. On my best day I am still going to miss something that another instructor would see. But if I have some rigor applied to, and some science of how I’m going to grade, breaking all those components down, then I feel that I grade in a very even fashion across all 16 students. And my number one worry as an instructor is that a student would think, and they could, they can appeal their grade, I’m not offended by them appealing their grade. I would feel wrong if they appeal the grade and someone could come in and look and say well, this exact same question you gave, you know, [one student] an 85, but you gave yourself a 92, and they say exactly the same thing. That’s embarrassing to me. I don’t ever want to be perceived of that, you can look at all of mine and I applied the same logic. It may be flawed logic, but I applied it evenly across the board.”

Dexter explained how he expected more out of certain officers, a norm-referenced approach. “My primary technique is; I attempt to the best of my abilities, to use inside of our courseware, the 1009s, writing, speaking, and communication [assessment forms]. I don’t do them as deliberately as I possibly could necessarily with where the students are supposed to judge their own responses and then, instructors are supposed to judge their own responses, but multiple sub-tasks. You know, I’m not going to put 2, 6, 10, 20, 30, 1’s, 2’s, 3’s, 4’s, and 5’s against each little individual element. I look at it in the aggregate or the whole, but I do use the material for supporting my ability to make an assessment on the student’s performance. I look critically at each individual officer’s responses to the question, but I, likely, either knowingly or unknowingly, introduce my own personal bias based on who the individual officer is, their background, their experience. So, I have different expectations for different officers, so I’m aware of that individual bias. I expect more out of maneuvers, fires, and affects officers

regarding Army tactical doctrine type requirements versus a non-standard, or specialty type officer like a doctor, a lawyer, army nurse, whatever it may be.”

Dexter continued with his discussion of norm-referencing, “What I do generally is, as I’ve learned over time, I read the questions of each individual officer’s response before I actually start to award or assign points or percentages for whatever their assessed performance is, because they’re, and I’m aware of the fact that the bell curve moves left and right so you read an answer to a question from an individual officer and you think it sounds either pretty good or pretty bad, but you’ll read the answer from an individual, a different individual officer, and you realize that the one you thought was really good wasn’t that good because this one’s better, or vice-versa. The one you thought was really bad wasn’t that bad, because the one you’re reading now is worse than the last one. So, it’s a floating judgment... I know we’re supposed to be judging against the standard of expected response to the question, but there is an introduced bias of comparison between individual officers that I’m responsible for and their individual performance against each other as well. Not just the question, but their individual, you know the other officers in their group. There is a comparison that’s just, it’s a reality.”

Dexter concluded, “But then [there is] another aspect of evaluation, tied to what I call the whole officer concept. ...sometimes it’s not necessarily important that they have a perfect answer to the question, it’s more particularly that they have demonstrated through their work, their answer to the question, whether or not they in fact committed to the appropriate, professional level of effort to try and respond to the assessment. Sometimes you can see that there’s an individual officer, or maybe one or two individual officers, but they’ll sort themselves out over time that [their] level of effort is clearly visible in some cases relevant to student performance. I will use that information then for a different ability to coach, mentor, and counsel, and provide feedback to the students as opposed to, did they learn what they were supposed to know, did they have the ability to share or communicate that back? It’s another aspect of an expectation of their performance.”

Jack started by describing criterion-referenced grading, “...most of them are metric in their construct. That’s generally the pattern I use. Actually I think all of them are. They’re given [a] certain number of points by section. If you want to look at commander’s intent, say for instance it’ll have X number of points for purpose, X number for key tasks, and then X number for end state and conditions. And that’s the pattern I use.”

Jack later described using a norm-referenced approach, “Does the instructor have a process to take unforeseen results from a summative assessment, and use it in a formative way later on in the course to address those areas of deficiency? I think instructors can vary greatly in what they focus on. The student experience level, could, does vary greatly within a classroom. You have officers who have already been in their KD billet and have a great deal of experience, and others just got promoted to major, and you have to, I won’t say treat those two different populations differently, but you do have to, when you’re taking the results of a summative assessment, and that latter group, with little experience, you do have to create some sort of formative platform for them to bring them to a higher level of learning.”

Arlen described stratification and a norm-referenced approach, “I think the one challenge in the college that maybe you didn’t ask is the question on the summative assessments is, a lot of in tactics is related to people who have been in the maneuver field and in your class you have a group of students that have been in the maneuver area of the military, and have some experience, and you have others in maybe logistics or inter-agency, that don’t, and the challenge that I have is, having that experience definitely will help them on the exam, and how do I differentiate between someone who’s from a foreign military, from a different service, from a different thing, and how do I measure their performance? Do I measure it based on one standard of, this is the maneuver standard for someone to have, this is the right answer because of all this experience they can put together a better package, or do I take into account where their background was when I’m doing a summative assessment? You learned a lot because you didn’t have the experience and you did well, but, really this guy who maybe didn’t learn very much but had the experience gave a better answer, and how do I relate those two together, and that’s always a challenge I have when the assessments of the students who maybe do not have a strong maneuver or tactical background. How do I assess them properly and fairly and not discount the stronger students who have the background, but maybe didn’t do as much research in the topic but did it because of the experience and those are hard to balance, and I haven’t, I haven’t quite, in the two years, figured out the right mix other than again that review of going back to looking at the assessments when they’re done and checking out the highs and lows and seeing if, you know, I can adjust, if I should adjust those to some extent because of the other circumstances, but that’s a challenge, and I don’t know, I haven’t figured out the right answer to that one yet.”

Hank stratified students, “I think there are some that, if it’s their background, they’re going to try to show their stuff. I think Infantry officers, specifically the maneuver guys, are going to try to show their stuff. There are some others that, just the way they are, they’re perfectionists, they’re going to. But then there’s others that, this is not their bread and butter, they’re probably not going to do this, and they’ll put minimal effort into it, so, yeah, I do think there’s different groups that will put more effort into it than others.”

Hank described grading on a curve. “I’m going to give a little bit of background first. You know, we have these great rubrics, and I know that they say, OK, here’s what an “A” is, here’s what a “B” is, here’s what a “C” is. At the same time, we were also briefed by the previous DC (Deputy Commandant), not everyone is above average. Okay, so, it’s kind of a balance between, and I don’t think we’d ever say we’re grading on a bell curve, but, it’s kind of a balance between we’re grading against the standard, versus, we’re grading to identify who are our very best. And so I think that’s the dichotomy that all faculty members have to face when they’re doing their assessments. So, what I do is, I generally go through all of them, and I will kind of put my comments and my assessments on them as to which one is the best, and which one is not so good, and kind of rack and stack them, and then I’ll go through them again, and I’ll look at the standards, at that point, you know, I’ve already kind of looked at the standards, I know when I’m going through there, looking for what’s right, what’s wrong, but then I’ll go through and really look hard, at that point I’m going to look at identifying who’s my top tier and assign them an, you know, an “A.” Then I’m going to take some others and I’ll look at them, and they may be pretty good, they might be, but they’re not above average. So what’s not above average to me? Not above to me is somewhere about a “B” plus, or an A- minus. And then you have those who are, you know, on the bottom end of average, and those are my strong, my B’s. So, I would say it’s a balance. But that’s kind of how I do all of my assessments. Whether it’s a written one, whether it’s the best, who provided the best answer? If they provided the best answer, I want to make sure that they get recognized. I don’t want to say, well here’s the standard, it wasn’t very high, everyone got an “A,” but yours was the very best. I just don’t think that is rewarding the person who put the most effort into it. So, we’re not saying it’s a bell curve, but, there are some that are clearly identified as better than others. Is that confusing enough?”

Hank concluded, “Well I also think that a “C” is not average. A “C” is failure here. So, average here is a “B.” And so, you say, well not everyone is above average. I don’t know what

percentage is above average. You say, well that's less than 50%? So, there's all sorts of things to think through as you go through this, but are you saying a "C," well a "C" means that you are below and you're failing and if you don't pick it up well then you're not going to pass. So, we're saying is this worth really failing? I know you call it a "U", but is it failing? Well, I think I can tweak them a little; I can work with this person and, uh, make sure that they don't fail so I give them a B-minus, or a "B."

Arlen discussed his grade curve, a part of norm-referencing. "I'll go through first and put a draft grade on every assessment, kind of review them and see where they are, and then I'll see how all the students fit in. So, I kind of try and look at what's the best and worst, and before assign a final grade, I'll go back and say hey, because this was the first one I graded, did I grade this one too hard, but I'll kind of try and re-look the outliers, probably the ones that are in the middle of the curve, they'll probably stay the same, but the highs and lows I'll say hey, was I too hard on this student, or I'll re-look some of their questions. If I took an excessive amount of points off, or gave them too high of a grade, I didn't give a lot of A pluses, I think they really had to be challenged, but I'd re-look those, and maybe either change it up or down depending on what I thought, but, go through and grade them all, and then I look at the highs and lows and see if I need to adjust them, and that seemed to work out okay."

The stratification of students during the grading process was typical among DTAC instructors during the grading process. Common summative assessments were graded in more of a nor-referenced approach than a criterion-referenced approach as designed. The unintended consequence of this was that student work was judged against their peers rather than the assessment criteria they used when taking the assessment. Some teams choose to calibrate their grading amongst team members.

Calibration amongst teaching team members

Five instructors described calibration within their teaching teams, possible because of common summative assessment used by all.

Archie described team calibration. "One thing I've done this year that I think has really helped is, I will grade my assessments and then as an after action review, I'll sit down with my three other partners on my teaching team and we'll compare across staff groups where the entire section struggled with a question and why. Was it the way we taught the lesson? Was it the way the question was framed for the student? Was it a blind spot in all the student's learning or

experience set that caused them to answer the question particularly poorly or particularly well, for that matter? So, that's one way, or that's one thing we've done. If I struggle with a particular student's answers a lot of times I'll set those papers to the side, but then I'll also ask for a second set of eyes. You know I haven't graded this one yet for final, but I'm struggling with this student's answer, I'll ask a fellow instructor, Can you look at this? Tell me what you think. Give me your thoughts to help kind of focus my grading. Because after you do sixteen pretty in-depth assessments, after you get to the sixteenth one, you're kind of burned out."

Allen described how calibration works within his team. "The whole calibration piece, definitely. ...the O300 COA Sketch and Statement is one that really stands out. As we all sat down first of all and looked at the rubric that came out for it, and then got through with our initial look at it, and said OK well here's some things that we can see already that are going to cause problems, so what do we need to do across the board within our teaching team, to try to get some commonality? Then going back and forth with the lesson author to say, OK, is this still within the bounds, you know, the left to right limits of what you're tracking for what you wanted? And then at the end of the block, going back to them and saying, OK, here's you know, the 28 or so extra, issues that we found with this particular, assessment and going through and explaining the reason why for all that piece of it there. Also, talking within the teaching team, ...trying to make sure that one of us isn't essentially being the candy man [easy grader] I found an extra 40 gigs or whatever that they could have picked out on that particular assessment. Well, you're really going outside the bounds on what you're giving them on this."

Allen "A lot of this year was me trying to get myself calibrated. I probably took a lot longer with doing my grading than most of the other folks on the grading team just because it was my first time going through a lot of these. First of all, [I] just go through everything, identify what I thought was incorrect, identify things that to me were maybe incorrect, was incorrect, or it wasn't highlighted in the rubric. Those are the things that I would go back and definitely talk to the other DTAC instructors and say, OK, were you counting off for something like this? Were you counting off for something like that? Those sorts of things there. Then, after going through and trying to lay all 16 assessments out and saying, ...basically here's the top pile, here's the middle pile, here's the bottom pile, and go back through again re-assessing everything again and saying, OK, these guys are clearly in the middle band, now let me figure out kind of where that middle band stands and what grade gets assigned to them, and the same for the top band, and

those guys who are at the bottom of the ladder on things. Then, especially the guys who were at the low end of the spectrum, and the guys who were at the very top end of the spectrum, going back and talking to the, to [instructor name deleted] or [instructor name deleted] a couple of the other experienced guys and say, OK, here's what I'm thinking, does this make sense to you? There's been a couple times when they've said, no, you're a little bit too harsh on this, or you're being a little bit too, forgiving on some of these; trying to get it all calibrated."

Alex explained comparing his grades with other members of his team. "One of the first things that I do is compare how my students graded out vs. my three other teaching peers within DTAC. For this year, that was... [Deleted part that ID'd staff groups by number] Were my guys and gals as associated with those? Because once you worked with your teaching team partners within the department, you know, OK, well I know [participant's name removed] because I share an office with him, things that he thinks are important that he stresses, and [instructor's name removed] teaches beside, OK, I know what they did so let me see if I sort of graded out about the same as them, then I feel that I had us on the right path."

Caleb explained different ways teams can calibrate. "It depends on the subject matter. And part of this is in terms of, and I not I'm mixing apples and oranges here, but a lot of it has to do with the team dynamics in terms of the opportunity, not necessarily the opportunities, but, there are ways that you can work with the other teammates. I generally will sit-in, and other guys sit-in on my class so that we can integrate and see where those weaknesses are in the group, and that's a good technique in terms of seeing where we may need to collectively adjust our teaching plan. But I've not had an issue going in, from my experience, going in another block, maybe for 10 minutes or during the tail end of SGA time or whatever it is, kind of going over that stuff. Perhaps a better way, or a more effective way, organizationally, would be to have, you know if you have, let's say you have three assessments in one week and you've got 10 days to [grade each] each, so, you know in two weeks, maybe another way is to at the beginning, you know, hey, let's review all these assessments we did. They can do it all at once. Now, there's, there's pros and cons, you know, with that, but there's different ways to do that but I haven't experienced any challenges on my own team by doing that."

Fred described how he progressed to participating in team calibration. "In a systemic way. I mean your first year of teaching, you have no idea [how to grade]. You've got the standard and you kind of apply it as best you know. But the second year you know enough about

it to start looking left and right, and say, okay, well what are the other guys [on the teaching team] thinking about this one. By year three or four, you have a pretty good sense of where you stand and how hard you need to be on things.

Summary

Semi-structured interviews with the DTAC instructors who participated in this study provided answers to the research questions. Key findings discovered during this research were supported by the literature review. It did not make a difference to instructors if an assessment was labeled summative or formative – instructors gave feedback to students when they could, in different ways. Instructors modified their teaching and learning activities after the use of common summative assessments. Instructors used many similar practices to prepare students for upcoming summative assessments. Instructors used many similar practices when implementing the grading process. Chapter Five summarizes the study, states conclusions based on the findings and the literature review, discusses implications for practitioners, and makes suggestions for future research.

Chapter 5 - Analysis, Discussion, Implications, and Recommendations

Introduction

The final chapter of this qualitative research case study on the formative use of common summative assessments during the U.S. Army Command and General Staff Officer's Course (CGSOC) is organized with: a summary of the study; the context and major findings; conclusions based on evaluation of the findings; a discussion of the implications for practitioners; and suggestions for future research.

Restatement of the Problem

Because assessment of student learning is context-dependent (Walvoord & Anderson, 2010), and CGSOC had a distinctive environment, previous literature did not fully address, from the higher education military instructor's perspective, the effect that using common graded assessment instruments had on the activities used to develop student learning. Instructors who teach standardized curriculums that use common summative assessments of learning need methods to develop student learning that are effective in their environment, where the lack of control over scheduling and curriculum limits the practices they can employ (Bloxham & Boyd, 2007). Results from research in other environments could not be generalized to the CGSOC environment.

Restatement of the Purpose

The purpose of this exploratory case study was to understand, from the Department of Army Tactics (DTAC) instructor's perspective, if there were formative ways common summative assessments were used to improve student learning and what similar practices, if any, instructors employed as a result of using common summative assessments of student learning.

Research Questions in Review

Two research questions guided this exploratory research on the formative use of common summative assessments:

Research Question One

Are there formative ways Department of Army Tactics instructors use common summative assessments?

Research Question Two

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

Discussion of the Study

Four conclusions were drawn from analysis of this research. First, the formative use of common summative assessments, especially feedback given to students, was typical of the Department of Army Tactics instructor, essentially a normal part of the assessment process. This important finding should be addressed in the DTAC curriculum. Second, DTAC instructors did not have a common understanding of the difference between summative and formative assessment, which mirrors findings of teachers at other learning institutions. How they used the information gathered was more important than what the assessment instrument was labeled. Third, “teaching to the curriculum” instead of “teaching the test” was typical across DTAC, an indication that the instructors saw their role beyond just preparing students for upcoming assessments. Fourth, the stratification of students during the grading process was typical, with the unintended consequence of students not being judged on quality of work alone. The analysis of each of these conclusions is discussed in detail below, aligned by research question. Research question one aligns with the first two major findings, and research question two aligns with the second two major findings. All findings convey the instructor’s perspective; students were not interviewed.

Research Question One

Are there formative ways Department of Army Tactics instructors use common summative assessments?

The formative use of common summative assessments, especially feedback given to students, was typical among Department of Army Tactics instructors.

The use of feedback to improve student learning is widely accepted as the key element of formative assessment. Taras (2009) asserted that most feedback in higher education comes from graded work and to not link formative assessment with summative assessment in this context risks losing “the most powerful and central learning tool [higher education] has” (p. 66). Taras’ assertion aligned with actions of the participants in this study who provided students feedback on their common summative assessments in written form, by analysis of common trends, and by referencing applicability to other assignments. Instructors provided written, individualized feedback to students on their common summative assessment papers as a standard practice. It did not make a difference to the instructor if an assessment was labeled summative or formative; feedback beyond a grade and the CGSC 1002 Assessment Form was provided to students. Although not as immediate as some definitions of formative assessment require, descriptive feedback intended to increase student learning was provided in enough time for instructors to adjust classroom behavior. Instructors discussed common errors, trends, and themes observed on graded common summative assessments with students as a group or individually, at times when the schedule allowed. Instructors provided students feedback that was applicable to assignments from other teaching departments, specifically in the area of communication in the form of writing.

DTAC instructors all engaged in some manner of the formative use of summative assessment, with feedback given to students being the most prevalent. Feedback in particular has been shown to help low achievers more than others (Black et al., 2003a; Harlen, 2004). Formative assessment is a powerful tool, with descriptive feedback its key characteristic. Based on Chappuis’ (2015) meta-analysis of the characteristics of effective feedback shown below, DTAC instructors gave effective feedback from the results of common summative assessments:

1. Directs attention to the intended learning, pointing out the strengths and offering specific information to guide improvement.
2. Occurs during learning, while there is still time to act on it.
3. Addresses partial understanding.
4. Does not do the thinking for the student.
5. Limits corrective action to the amount of advice the student can act on. (p.95)

The formative use of summative assessments to increase student learning was typical among DTAC instructors. Similarly, DTAC instructors were not concerned with what assessments were called, the use of the information gathered to increase student learning was more important than how an assessment was labeled.

DTAC instructors did not have a common understanding of the difference between summative and formative assessment.

There is not a universal definition or understanding of summative or formative assessments in the field of education mirroring DTAC instructors who also did not have a common definition or understanding. Suskie (2009) noted that, “Because the assessment of student learning in higher education is relatively new compared to many other fields of study, and because it has been undertaken by people from disciplines with widely differing orientations, the vocabulary of assessment is not yet standardized” (p. 3). Assessments can and should be designed to accommodate both summative and formative purposes. As shown in this research, the unintended use of a balanced approach between summative and formative assessment purposes made the best use of valuable time. CGSOC is a higher education program where students attend classes daily and are in class at least four hours each day. To not plan for assessments with balanced purposes would be a wasted opportunity. Harlen’s (2012) “dimensions of assessment” chart (Figure 2.4 of this study) provides a more relevant way to think of assessments used by DTAC instructors. Harlen (2015) used this figure to illustrate how the “blurred distinction between assessment to help learning (formative) and assessment to report learning (summative) indicates that the relationship between formative and summative assessment might be better described as a dimension rather than a dichotomy” (p. 98). DTAC

instructor practices covered the entire range of dimensions or spectrum with common assessments labeled “summative.”

Chappuis (2015) concluded that there was a common thread woven throughout formative assessment research, articles, and books: “It is *not the instrument* that is formative; it is *the use of the information* gathered, by whatever means, *to adjust teaching and learning*, that merits the “formative” label (p.4-5). Correspondingly, DTAC instructors were not concerned that the common assessments they gave their students were labeled summative; they were concerned with using the information they gathered from the assessment in a formative way. How they used the information gathered was more important than what the instrument was called.

The timing and timeliness of the feedback provided to students and adjustments to teaching and learning activities are areas where DTAC instructor’s actions differed most from “timing” definitions of formative assessment. The essence of Shepard’s (2008) description of the timing, used by other educators, worded differently, but with the same characterization, “Formative assessment is defined as assessment carried out during the instructional process for the purpose of improving teaching or learning. ...What makes formative assessment formative is that it is immediately used to make adjustments so as to form new learning” (p. 281). “Immediately used to make adjustments,” does not describe DTAC instructors’ actions. The claim that feedback must be immediate to be formative does not agree with the findings of this research. The findings correspond with Martin and Collins (2011), who highlighted the difference between those who use time as the determining factor between formative and summative assessment and those who use the purpose of the assessment as the determining factor. They used the example, “A teacher could give an examination at the end of the semester which is for the purpose of assigning grades (summative) and then use the data gathered from that examination to improve her curriculum for the next semester and turn the summative assessment into a formative assessment” (p.132). During this research, participants understood and were comfortable with the reality that some gains will not follow immediately.

While giving feedback and not having a common understanding of assessment terms were common to DTAC instructors, these findings were focused on how summative assessments were used formatively. The following section transitions from a formative focus to similar practices instructors used as a result of common summative assessments and addresses findings in the areas of teaching and grading techniques.

Research Question Two

Are there similar practices used by Department of Army Tactics instructors as a result of using common summative assessments? If so, what are they?

“Teaching to the curriculum” instead of “teaching the test” was the norm in DTAC, an indication that instructors saw their role beyond just preparing students for assessments.

While preparing students for upcoming summative assessments, “Teaching to the test” was common among DTAC instructors, whereas “teaching the test” was not. These techniques are better described in terms Popham (2001) used, “curriculum teaching” and “item-teaching.” Teaching the test, or item teaching is considered unethical because it misrepresents how much students really have learned about a topic and results in unreliable data being used to make decisions at higher levels. Item-teaching is linked with “high-stakes” tests, like those used to evaluate the effectiveness of a school’s teachers. These type tests are graded by someone other than the student’s instructor. Common Summative Assessments at CGSOC are graded by the student’s instructor, but could be considered high-stakes because of the negative effect failing could have on a student officer’s career.

Preparing students for upcoming assessments included holding practice sessions focused on the type of topics/skills students would see on the common summative assessment. Students were shown examples of “what right looks like,” or “a” solution before taking summative assessments to show them what will be expected of them. The use of curriculum-teaching by participant instructors who were preparing their students to take common summative assessments was not because of a written rule or policy, it was because instructors felt that while it is important for students to pass their assessments, it is perhaps more important to focus on teaching problem-solving and improvement of written communication skills, better preparing students for their next assignments. “Why” curriculum-teaching is important needs to be understood and the potential for and causes of misuse identified and mitigated.

The use of the test’s actual test items for practice or items so similar that they are almost indistinguishable from test questions is item-teaching, and this results in deceptive outcomes. If an instructor uses item-teaching and the skills for a particular block of instruction that CGSOC students are expected to master are not all represented on the assessment, then the results will not accurately represent student mastery. An instructor who chooses to item-teach is teaching a

sample of skills chosen to represent a larger number of skills, and the outcome will be treated as representative of the student's mastery. A perfect score would infer that a student can satisfactorily do problems of same the type but in a different context, which may not be true (Popham, 2001). This misrepresents how much a student really knows about a topic. When decisions are made at the program level based on this information, resources, usually in the form of instructor/student "contact hours" will be incorrectly allocated by higher administration.

The use of broad learning objectives in curriculum design can lead to more item-teaching than to the learning objective, if assessments are not aligned with the curriculum. Resnick and Zurawsky (2005), contended that "When teachers match their teaching to what they expect to appear on state tests of this sort, students are likely to experience far more facts and routines than conceptual understanding and problem-solving in their curriculum....Narrow tests may not serve simply as the floor, but can become the de facto curriculum" (p. 11). Instructors who have their students prepare specifically for test questions are using time that could be used to teach the full curriculum. Modifying teaching and learning activities are widely accepted as key elements of formative assessment. Participants in this study modified or adjusted their teaching and learning activities specifically because of common summative assessments. Participants described adjusting learning activities or teaching style prior to and after the assessment was given. Unlike many definitions of formative feedback, which require immediate classroom activities in response to gaps recognized during student learning, participant adjustments were often weeks later, with the understanding that some gains won't follow immediately. Additionally, participants made recommendations to curriculum developers on how to improve future common summative assessment instruments.

Another common experience with the use common summative assessments among DTAC instructors was the practice of stratifying students during the grading process. Again, the instructors' goal was to improve student learning.

The stratification of students during the grading process was typical, perhaps with unintended consequences.

The intentionally diverse population of each instructor's group of 16 students led to common summative assessments being graded in more of a norm-referenced approach than a criterion-referenced approach, as they were designed to be graded. Bloxham and Boyd (2007)

contended that differences in instructor approaches to grading assessments, categorized as norm-referenced or criterion-referenced, often result in unintended outcomes.

Norm-referenced assessment is designed to distribute student performance over a range, for the purpose of discriminating between those who do well and those who do less well (Bowden & Martin, 1998). An example of this would be if 30% receive an A, 60% a B and the remaining 10% a C. Criterion-referenced assessment judges students against a set of criteria that is linked to desired learning outcomes. Bloxham and Boyd (2007) maintained that “whereas with criterion-referenced assessment all students have an opportunity to do equally well, a norm-referenced approach will almost always create a distribution of grades” (p. 82). Educators feel that norm-referencing is still being used because it is easier and some grading schemes can only be understood and applied by referencing student work (Price, 2005).

What most participants in this study described was their stratification of students based on student experience and/or motivation; what was essentially a norm-based grading approach although CGSOC summative assessments were designed to be criterion referenced. Price (2005) argued that because it is fairer to students, criterion-referenced assessment is generally considered the better of the two approaches. Students deserve to know assessment criteria beforehand and be judged based on the quality of their work rather than have their performance compared to their classmates. Bloxham and Boyd (2007) concluded that applying standards is not straight forward and requires contextual interpretation. Instructors in higher education use a combination of judgment and the application of grading standards criteria which are greatly influenced by the norms of the institution. The norm-based approach was typical of DTAC instructors, out of a desire to provide feedback focused on the student’s likelihood of having to use the skill being assessed in the future. In the past student test ID numbers were issued to all CGSC students, and students wrote these numbers on the name line instead of their name, allowing instructors to grade their work anonymously. Instructors would not know whose assessment they were grading until they were entering the grade. The issuing of student ID numbers was discontinued but may want to be reconsidered as an option for individual instructors who would like to evaluate their own grading process.

Instructors felt that students were motivated by different things, and that prioritization because of multiple conflicting requirements was the cause of many not doing well on assessments, especially take-home assessments. Instructor understanding of different student

motivations and approaches to assessments may assist with the recognition of unhelpful approaches to study and prompt coaching designed to increase student learning. Chappuis (2015) discussed goal orientations as the key to understanding different student motivations answering the question, “Why am I doing this assignment?” (p.15) Student goal orientations fall into three categories (Ames, 1992; Schunk, 1996; Black & Wiliam, 1998a; Halvorson, 2012):

1. *A learning orientation*, where the goal is to get better. When faced with difficulty, they get more involved. Motivated to learn and a willingness to engage in the process of learning. Have an intrinsic value of learning. They tend to seek help frequently while developing competence and then avoid help once mastery is perceived.
2. *A performance or ego orientation*, where the goal is to prove ability or hide a perceived lack of ability. Want the recognition of others and to be seen as smart. Achieving success with least effort is their methodology, leading to reluctance to engage in effort-based learning activities. When faced with difficulty, they exhibit anxiety and poor performance, sometimes leading to cheating. They tend to avoid seeking help to hide their perceived lack of ability.
3. *A task-completion orientation*, where the goal is to get it done and get a grade. Not interested in learning and mastery. Will expend just enough effort to get assessment turned in. When faced with difficulty, looks for ways to get completion, not understanding.

Stratification of students during the grading process was common during the grading process. Students were often graded as if the assessments were norm-referenced when assessments were designed to be criterion-referenced because of the purposely diverse student population. Calibration amongst teaching team members was enabled by the use of a common assessment instrument; this was especially helpful to newer, less experienced instructors. Instructors used the term prioritization vs. motivation to describe why some students did not do as well as others on summative assessments which were mostly take-home. These findings should inform other instructors who find that they are facing similar situations.

Based on an analysis of the major findings, the following implications are suggested. Implications are grouped in two sections: Military Education and Adult Education.

Implications for Military Education

- The formative ways feedback from common summative assessments was used by participant instructors could be included in the DTAC curriculum to increase student learning throughout the student population.
- The use of combined summative and formative assessment, for different purposes, could be recognized as an opportunity to make better use of available time.
- The practice of teaching to the curriculum used by instructors to prepare students for assessments and increase student learning could become part of the DTAC faculty development program.
- The practice of instructor stratification of students during the grading process can be better understood by those who use assessment data to make decisions about future curriculum design and the allocation of contact hours.
- Findings /conclusions of this research could be generalized to similar curriculums that are taught in a military-like environment.

Implications for Adult Education

- This case study can add to the body of knowledge on the formative use of summative assessments in higher education by the discussion and example of how participant instructors used common summative assessments in multiple formative ways.
- This case study can add to the body of knowledge of the formative use of summative assessment by the discussion and example of the timing of feedback “in time to act” in the context of higher education.
- This case study may add to the body of knowledge of the definition of assessments by its discussion and example of the “blurred distinction” between formative and summative assessments in higher education.

Recommendations

- Balance the use of summative and formative purposes throughout the DTAC curriculum by the use of embedded assessments. This would ensure that the power of formative assessment is not an opportunity missed.

- Include a table with each assessment that describes its multiple purposes to instructors.
- Provide students with not only strong, but also weak examples of products that they will be required to develop as part of assessments, giving them an understanding of the full range of possible responses.
- Make available student assessment identification numbers for use by instructors who want to grade anonymously with the goal of eliminating grading bias.
- Provide timely feedback to students beyond just grades by developing assessments designed to be timely.

Recommendations for Future Research

- Students are the consumers of assessment information. Looking at assessments from the student's point of view would provide insights into what motivates them, how useful feedback given to them was, including timeliness, how classroom activities prepared them for assessments, and the student's role in assessment.
- Analyzing instructor written comments given to students on assessments could provide a deeper understanding of the instructor's role in the learning process. Specifically the use of descriptive versus less desirable evaluative feedback, and how students perceive the usefulness of provided comments.

Reflections

Looking back at the entries made in my reflective journal throughout the research process, it was clear that I have come full circle in my thinking about assessments. When I began, I was obsessed with defining summative and formative assessments each as a separate instrument, each with fixed characteristics. The literature on assessments contributed to this obsession, because most of it sought to define assessments in finite terms as well. Now, at the end of my research, I find assessments to be measurable, but along a dimension or on a spectrum and very much contextual in nature. The same assessment could be used by an instructor to obtain feedback identifying needed adjustments to teaching and provide written feedback to

students, and it could be used by curriculum developers and decision makers to make program level decisions about resources such as the allocation of teaching hours for particular skills. I am comfortable with an assessment being used for multiple purposes, even if it was not designed that way when the focus is on increasing student learning.

What surprised me most, and led to my current view on assessments was that it did not appear to matter to participants what an assessment was labeled, they were focused on giving feedback and increasing student learning. Participants found the discussion of assessment definitions professionally interesting, but not important enough to consider changing their actions because of a definition, they were focused on giving feedback to students on how to improve and to curriculum developers about how to increase student learning, from their perspective. Education is contextual in nature and involves many professionals each trying to do their best to help students learn, there is opportunity and power in combining multiple levels of purpose in the same assessment to obtain feedback that can be used by many.

Concluding Remarks

This research confirmed that the formative use of summative assessments was typical among Department of Army tactics instructors and that continued or expanded formative use of summative assessments will increase student learning. Because so much of assessment is context dependent, this research will add to the body of knowledge in a particular area that the current literature did not fully address; the formative use of common summative assessments in higher education. Instructors and curriculum developers in the Department of Army Tactics and other departments in the Command and General Staff College will be able to use the findings of this research to inform their teaching styles and techniques and improve student learning.

References

- Abbott, D., Broadfoot, P., Croll, P., Osborn, M., & Pollard, A. (1994). Some sink, some float: National curriculum assessment and accountability. *British Educational Research Journal*, 20: 97-117
- Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. Thousand Oaks, CA: Corwin Press.
- Allal, L., & Lopez, L.M. (2005). Formative assessment of learning: A review of publications in French. In Office of Economic Co-Operation and Development, *Formative assessment: Improving learning in secondary classrooms*. Paris: OECD Publishing.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology*, 84(3), 261-271.
- Angelo, T. (1995). Reassessing (and redefining) assessment. *AAHE Bulletin*, 48 (3), 7-9.
- Banta, T., Lund, J., Black, K. & Oblander, F., (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.
- Banta, T., Jones, E., & Black, K., (2009). *Designing effective assessment: Principles and profiles of good practice*. San Francisco: Jossey-Bass.
- Bell, B., & Cowie, B. (1997). Formative assessment and science education. Research report of the Learning in Science Project (Assessment). Hamilton, New Zealand: University of Waikato.
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85: 536-553.
- Bennett, S., & Wragg, E. (1992). A longitudinal study of primary teacher's perceived competence in, and concerns about National Curriculum implementation. *Research Papers in Education*, 7 (10): 53-78.
- Biggs, J. (2003). *Teaching for quality learning at university*. Buckingham: The Open University Press.
- Black, P.J. (1993). Formative and summative assessment. *Studies in Science Education*, 21 (1): 49-97.
- Black, P.J. (1998). *Testing, friend or foe?: the theory and practice of assessment and testing*. Psychology Press.
- Black, P.J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5 (1), 7-73.
- Black, P.J., & Wiliam, D., (1998b). Inside the black box: raising standards through classroom assessment, *Phi Delta Kappan*, 80 (2).

- Black, P.J., & Wiliam, D., (2003). "In praise of educational research: Formative assessment." *British Educational Research Journal*. 29 (5): 623-637.
- Black, P.J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). *Working inside the black box: assessment for learning in the classroom*, Phi Delta Kappan, 86 (1), 9-21.
- Black, P.J., & Wiliam, D., (2009). Developing the theory of formative assessment. *Educational assessment, evaluation and accountability*. 21(1): 5-31.
- Black, P.J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D., (2003a). *Assessment for learning: Putting it into practice*. New York: Open University Press.
- Black, P.J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D., (2003b). The nature and value of formative assessment for learning. Paper presented at AERA Chicago 22 April, <http://www.kcl.ac.k//depsta/education/hpages/pblavkpubs.html>.
- Black, P.J., Harrison, C., Lee, C., Marshall, B., & Wiliam, D., (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan*. 86 (1) 8-21.
- Bloom, B. (1969). Some theoretical issues relating to educational evaluation. In R.W. Tyler (ed), *Educational evaluation: New roles, new means: the 63rd yearbook of the National Society for the Study of Education (part II)*. 69 (2): 26-50. Chicago, IL: University of Chicago Press.
- Bloom, B., Hastings, J. & Madus, G. (eds.) (1971). *Handbook on the formative and summative evaluation of student learning*. New York: McGraw-Hill.
- Bloxham, S., & Boyd, P. (2007). *Developing effective assessment in higher education: A practical guide*. New York: Open University Press.
- Bogden, R., & Biklen, S. (2007). *Qualitative research for education: An introduction to theories and methods*. Boston: Pearson.
- Bowden, J., & Martin, F. (1998). *The university of learning: Beyond quality and competence in higher education*. London: Kogan Price.
- Bredo, E. (1997). The social construction of learning, in Phye, G. (ed.) *Handbook of academic learning: Construction of knowledge*. San Diego, CA: Academic Press.
- Broadfoot, (2008). Assessment for learners: Assessment literacy and the development of learning power. In: Havnes, A. & McDowell, L. (eds.). *Balancing Dilemmas in Assessment and Learning in Contemporary Education*. New York, London: Routledge, 213-224.
- Broadfoot, P., & Black, P. (2004). Refining assessment? The first ten years of assessment. *Assessment in Education*. 11 (1): 7-26.
- Brookhart, S.M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*. 106 (3): 429-458.

- Brookhart, S. & DeVoge, J. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education*. 12: 409-425.
- Brown, S., Race, P. & Smith, B. (1996) *500 tips on assessment*. London: Kogan Page.
- Brown, S. (2004-2005) Assessment for learning. *Learning and Teaching in Higher Education*, 1, 81-89.
- Bullock, K., Bishop, K., Martin, S., & Reid, A. (2002). Learning from coursework in English and geography. *Cambridge Journal of Education*. 32: 325-340.
- Burke, K., (2010). *Balanced assessment: From formative to summative*. Bloomington, IN: Solution Tree Press.
- Butler, S., & McMunn, N. (2006) *A teacher's guide to classroom assessment: Understanding and using assessment to improve student learning*. San Francisco, CA: Jossey-Bass
- Caffarella, R. (2002). *Planning programs for adult learners* (2nd ed.). San Francisco, CA: John Wiley & Sons.
- Careless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*. 31 (2): 219-233.
- Carter, C. (1997/8). Assessment: Shifting the responsibility. *Journal of Secondary Gifted Education*. 68: 68-75.
- Chappuis, J. (2015). *Seven strategies of assessment for learning* (2nd ed.). NJ: Pearson Education.
- Chairman of the Joint Chiefs of Staff (2009), *Instruction 1800.01D. Officer professional military education policy*. 15 July 2009. Office of the Chairman of the Joint Chiefs of Staff, Washington, DC.
- Chairman of the Joint Chiefs of Staff (2012). *Joint education white paper*. Office of the Chairman of the Joint Chiefs of Staff, Washington, DC.
- Chickering, A. & Gamson, Z. (1987). Seven principles for good practice in undergraduate education. *American Association for Higher Education Bulletin*. 39 (7): 3-7.
- Coladarci, T. (1986). *Teacher efficacy and school climate*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Cowie, B., & Bell, B. (1999). A model of formative assessment in science education. *Assessment in Education*. 6 (1), 102-116.
- Creswell, J.W. (2007). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage Publications.

- Creswell, J.W. (2008). *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*, Upper Saddle River (3rd ed.), NJ: Pearson Education.
- Creswell, J.W.(2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage Publications.
- Crisp, B. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work. *Assessment & Evaluation in Higher Education*. 32: 571-581.
- Crooks, T. (1988). The impact of classroom evaluation practices on students. *Review of educational research*. 58 (4) 438-481.
- Davis, G., & Martin, J. (2012). Developing leaders to adapt and dominate for the army of today and tomorrow. *Military Review*: Sept-Oct 2012.
- Duncan, N. (2007). "Feed-forward": improving student's use of tutor's comments. *Assessment & Evaluation in Higher Education*, 32(3), 271-283.
- Dunn, K., & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment Research & Evaluation*. 14 (7) 1-11.
- Elton, L., & Johnston, B. (2002). *Assessment in universities: A critical review of research*. York: Higher Education Academy.
- Elton, L., & Laurillard, D. (1979). Trends in research on student learning. *Studies in Higher Education*. 4, 87-102.
- Ekman, R., & Pelletier, S. (2008). Assessing student learning: A work in progress. *Change*. 40 (4), 14-19.
- Entwistle, N. (1997). Contrasting perspectives on learning. In Marton, F., Hounsel, D., & Entwistle (eds), *The experience of learning: Implications for teaching and studying in higher education*. Edinburgh: Scottish Academic Press.
- Ewell, P. (2002). *Perpetual movement: Assessment after twenty years*. Keynote address at the American Association for Higher Learning Annual Meeting, Boston, MA. Reprinted in the 2010 Assessment Institute proceedings.
- Ewell, P. (2008). *U.S. accreditation and the future of quality assurance; a tenth anniversary report from the council for higher education accreditation*. Washington, D.C.: Council for Higher Education Accreditation.
- Ewell, P. (2010). *Assessment, accountability, and improvement*. National Institute for Outcomes Assessment, Occasional Paper #1.

- Flex, R., Cumbo, K., Borko, H. Mayfield, V., & Marion, S. (1995). *How “messing about” with performance assessment in mathematics affects what happens in classrooms*. (Technical Report 396). Los Angeles: Centre for Research on Evaluation, Standards and Students Testing (CRESST).
- Fraenkel, J., & Wallen, N. (2009). *How to design and evaluate research in education* (7th ed.). New York: McGraw-Hill.
- Gailbraith, M., Sisco, B., & Guglielmino, L. (2001). *Administering successful programs for adults: promoting excellence in adult, community, and continuing education*. Malabar, FL: Krieger.
- Garrison, C., & Ehringhaus, M. (2007) Formative and summative assessments in the classroom. Retrieved 8 July, 2011 from: http://www.nmsa.org/portals/0/pdf/publications/Web_exclusive/Formative_Summative_Assessments.pdf
- Gibbs, G. (1988). *Learning by doing: a guide to teaching and learning methods*. Oxford Polytechnic: Oxford.
- Gibbs, G., & Simpson, C. (2004-2005). Conditions under which assessment supports students’ learning. *Learning and Teaching in Higher Education*, 1, 3-31.
- Gipps, C., & Clarke, S. (1998). Monitoring consistency in teacher assessment and the impact of SCAA’s guidance materials at Key Stages 1, 2, and 3. Final Report, London: QCA.
- Glesne, C. & Pushkin, A. (1992). *Becoming qualitative researchers: An introduction*. White Plains, NY: Longman.
- Greeno, J., Pearson, P., & Schoenfeld, A. (1996). *Implications for NAEP of research on learning and cognition. Report of a study commissioned by the national academy of education*. Panel on the NAEP Trail State Assessment, conducted by the Institute for Research on Learning. Stanford, CA: National Academy of Education.
- Gordon & Rees (1997). High stakes testing: Worth the price? *Journal of School Leadership*. 7: 345-368.
- Guba, E., & Lincoln, Y. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco: Jossey-Bass.
- Guba, E., & Lincoln, Y. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage.
- Guba, E., & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage
- Hall, K., Webber, B., Varley, S., Young, V., & Dorman, P. (1997). A study of teacher assessment at key stage 1. *Cambridge Journal of Education*. 27, 107-122.

- Hall, K., & Harding, A. (2002). Level descriptions and teacher assessment in England: Towards a community of assessment practice. *Educational Research*. 44, 1-15.
- Halvorson, H. G. (2012). *Succeed: How we reach our goals*. New York, NY: Penguin.
- Hargreaves, A., Earl, L., & Ryan, J. (1996). *Schooling for change: Reinventing education for early adolescents*. New York: Falmer Press.
- Harlen, W., & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in education: Principles, policy and practice*. 4 (3), 365-379.
- Harlen, W. (1998). Classroom assessment: A dimension of purposes and procedures. *SAMEpapers*, 75-97.
- Harlen, W. (2003). *Enhancing inquiry through formative assessment*. San Francisco, CA: Exploratorium.
- Harlen, W. (2004a) A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Harlen, W. (2004b). A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Harlen, W. (2005). Teacher's summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16 (2), 207-223.
- Harlen, W. (2012). On relationships between assessment for formative and summative purposes, in Gardner, J. (ed). *Assessment and learning* (2nd ed.). Los Angeles: Sage.
- Harlen, W., & Crick, D. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning. In: *Research Evidence in Education Library*. London: EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Hartley, P. Skelton, A., & Higgins, R. (2002). The conscientious consumer: reconsidering the role of assessment feedback in student learning. *Studies in higher education*. 27(1): 53-64.
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Education Research*. 77 (1): 81-112.
- Heibert, E., & Davinroy, K. (1993). Dilemmas and issues in implementing classroom-based assessment for literacy (Technical Report 365), Los Angeles, Centre for Research on

Evaluation, Standards and Student Testing (CRESST), accessed from:
<http://www.cse.ucla.edu/CRESST/Reports/TECH365.PDF>

- Heritage, M. (2010). Formative Assessment and Next-Generation Assessment Systems: Are We Losing an Opportunity? *Council of Chief State Officers*.
- Herman, J.L., Osmundson, E., & Silver, D. (2010). Capturing Quality in Formative Assessment Practice: Measurement Challenges. CRESST Report 770. *National Center for Research on Evaluation, Standards, and Student Testing (CRESST)*.
- Hounsell, D. (2003). Student feedback, learning and development, in Slowery, M. & Watson, D. (eds) *Higher education and the lifecourse*. Buckingham: SHRE/Open University Press.
- Hu, S. (2005). Beyond grade inflation: Grading problems in higher education. *ASHE Higher Education Report*, 30, (6).
- Iredale, C. (1990). Pupil's attitudes towards GASP (Graded Assessments in Science Project). *School Science Review*. 72: 133-137.
- James, D., & Fleming, S. (2004-2005). Agreement in student performance in assessment. *Learning and Teaching in Higher Education*, 1, 32-50.
- James, M. (2008). Assessment and learning, in Swaffield, S., *Unlocking assessment: Understanding for reflection and application*. Abingdon: Routledge pp. 20-35.
- James, M., & Lewis, J. (2012). Assessment in harmony with our understanding of learning: problems and possibilities. *Assessment and learning* 2, 187-205.
- James, R. (2003). Academic standards and the assessment of student learning: some current issues in Australian higher education. *Tertiary Education and Management* 9 (3): 187-198.
- James, R., McInnis, C., & Devlin, M. (2002a). Assessing learning in Australian universities. Available at unimelb.edu.au/assessing-learning/ (assessed on 7 June 2014).
- Jerald, C. (2006). *Teach to the test? Just say no*. The Center for Comprehensive School Reform and Improvement. Washington, DC. Available at <http://www.readingrockets.org/article/teach-test-just-say-no> (assessed on 17 June 2016).
- Johnston, P., Afflerbach, P. & Weiss, P. (1993). Teacher's assessments of the teaching and learning of literacy. *Educational Assessment*. 1: 91-117.
- Keeton, M., Sheckley, B., & Griggs, J. (2002). *Effectiveness and efficiency in higher education for adults: A guide for fostering learning*. Dubuque, IA: Kendall/Hunt.
- Knight, P.T. (ed) (1995). *Assessment for learning in higher education*. London: Kogan Page Limited.

- Knight, P.T. (2000). The value of a programme-wide approach to assessment. *Assessment & Evaluation in Higher Education*. 25 (3), 237-251.
- Knight, P.T. (2002). Summative assessment in higher education: Practices in disarray. *Studies in Higher Education* 27 (3), 275-286.
- Kortez, D., Stecher, B., Klein, S., & McCaffrey, D. (1994). The Vermont portfolio assessment program: Findings and implications. *Educational Measurement: Issues and Practice*. 13: 5-16.
- Laurillard, D. (1997). *Applying systems thinking to higher education*. Milton Keynes: Open University.
- Laurillard, D. (2002). *Rethinking university teaching: A conversational framework for the effective use of learning technologies* (2nd ed.). London: Routledge Falmer.
- Light, G., & Cox, R. (2001). Assessing student assessment. *Learning and Teaching in Higher Education: The Reflective Practitioner*. London: Paul Chapman Publishing.
- Linn, (2000). Assessments and accountability. *Educational Researcher*. 29: 4-16.
- Looney, J. (2011). Integrating formative and summative assessment: progress towards a seamless system? *OECD Education Working Papers No. 58*, OECD Publishing.
- MacNaughton, G. (2001). Action research. In G. MacNaughton, S. A. Rolfe, & I. Siraj-Blatchford (Eds.), *Doing early childhood research: International perspectives on theory and practice* (pp. 208-223). Crows Nest NSW: Allen & Unwin.
- Marshall, C. & Rossman, G. (1998). *Designing qualitative research*. Thousand Oaks, CA: Sage.
- Marshall, C. & Rossman, G. (2006). *Designing qualitative research* (4th ed.). Thousand Oaks, CA: Sage.
- Marshall, C. & Rossman, G. (2011). *Designing qualitative research* (5th ed.). Thousand Oaks, CA: Sage.
- Martin, J., & Collins, R.A. (2011). Assessing and evaluating adult learning in career and technical education, in Wang, V.X. (ed.). *Formative and summative evaluation in the assessment of adult learning*. Hershey, PA: IGI Global.
- Marton, F., & Saljo, R. (1997). Approaches to learning. In Marton, F., Hounsel, D., & Entwistle (eds), *The experience of learning: Implications for teaching and studying in higher education*. Edinburgh: Scottish Academic Press.
- Marzano, R. (2010) *Formative assessment and standards-based grading*. Bloomington, IN: Marzano Research Laboratory.

- Maxwell, G. (2004) Progressive assessment for learning and certification: some lessons from school-based assessment in Queensland. Paper presented at the Third Conference of the Association of Commonwealth Examination and Assessment Boards; Nadi, Fiji.
- McCallum, B., & McAlister, S. (1993). Teacher assessment at Key Stage One. *Research Papers in Education*. 8(3): 305 - 328.
- McPartland, J. (1987). Changing grading and testing practices to improve student motivation and teacher-student relationships: Designs for research to evaluate new ideas for departmental exams and progress grades. Paper presented at the annual meeting of the American Educational Research Association, April, Washington, DC.
- Merriam, S. (1998). *Qualitative research and case study applications in education*. (2nd ed.). San Francisco: Jossey-Bass.
- Merriam, S., Caffarella, R., & Baumgartner, L. (2007). *Learning in adulthood: A comprehensive guide* (3rd ed.). San Francisco: Jossey-Bass.
- Miller, G. (1976). Continuous assessment. *Medical education*. 10, 81-86.
- Mruck, K., & Breuer, F. (2003, May). Subjectivity and reflexivity in qualitative research- The FQS issues. *Forum Qualitative Sozialforschung*, 4(2).
- Newton, P. (2007). Clarifying the purposes of educational assessment. *Assessment in education*. 14 (2) 149-170.
- Orsmond, P., & Merry, S., & Reiling, K. (2005). Biology students' utilisation of tutors' formative feedback: A qualitative interview study. *Assessment & Evaluation in Higher Education*. 30: 369-386.
- Orsmond, P., Merry, S. (2011). Feedback alignment: Effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*. 36(2): 125-136.
- Ortlipp, M. (2008). Keeping and using reflective journals in the qualitative research process. *The Qualitative Report*, 13(4), 695-605.
- Patton, M., "Quality in Qualitative Research: Methodological Principals and Recent Developments." Invited address to Division J of the American Educational Research Association, Chicago, April 1985.
- Patton, M. (1990). *Qualitative evaluation and research methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Patton, M. (1998). Discovering process use. *Evaluation*. Thousand Oaks, CA: Sage.

- Pellegrino, J., Chudowsky, N., & Glaser, R. (eds) (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Popham, W.J., (2001) Teaching to the test? *Educational Leadership*. 58(6): 16-20.
- Price, M. (2005). Assessment standards: the role of communities of practice and the scholarship of assessment. *Assessment & Evaluation in Higher Education*, 30(3), 215-230.
- Prosser, M., & Trigwell, K. (1999). *Understanding learning and teaching: the experience in higher education*. Buckingham: Open University Press.
- Race, P. (1993). *Never mind the teaching--feel the learning*. Staff and Educational Development Association (SEDA) London, Paper 80.
- Race, P. (1995). The art of assessing 1, *New Atlantic*, Autumn, 3-6.
- Ramaprasad, A. (1983). On the definition of feedback. *Behavioral Science*, 28 (1), 4-13.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Ramsden, P. (2003). *Learning to teach in higher education* (2nd ed). London: Routledge Falmer.
- Reeves, D. (2011). *Elements of grading: A guide to effective practice*. Bloomington, IN: Solution Tree Press.
- Resnick, L., & Zurawsky, C. (2005). Getting back on course: Standards-based reform and accountability. *American Educator*. Spring.
- Rowntree, D. (1985). *Developing courses for students*. London: Routledge.
- Rowntree, D. (1987). *Assessing students: how shall we know them?* London: Kogan Page.
- Rust, C. (2002). The impact of assessment on student learning: How can the research literature practically help to inform the development of departmental assessment strategies and learner-centered assessment practices? *Active learning in higher education*. 3(2): 145-158.
- Sadler, D.R. (1983). Evaluation and the improvement of academic learning. *Journal of higher education*. 54, 60-79.
- Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 145-165.
- Sadler, D.R. (1998). Formative assessment: revisiting the territory. *Assessment in education*. 5 (1) 77-84.
- Sadler, D.R. (2005). Interpretations of criteria-based assessments and grading in higher education. *Assessment and evaluation in higher education*. 30(2), 175-194.

- Schunk, D. (1996). Goal and self-evaluative influences during children's cognitive skill learning. *American Educational Journal*, 33(2), 359-382.
- Scriven, M. (1967). *The methodology of evaluation*. Washington, DC: American Educational Research Association.
- Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park, CA: Sage Publications.
- Scriven, M. (2006). Key evaluation checklist (KEC). Retrieved from http://www.wmich.edu/evalctr/checklist/kec_june06.pdf.
- Shavelson, R. (2006). On the integration of formative assessment in teaching and learning: Implications for new pathways in teacher education. In Oser, F., Achtenhagen, F., Renold, U. (eds.), *Competence-Oriented Teacher Training: Old research Demands and New Pathways*. Utrecht, Netherlands: Sense Publishers.
- Shepard, L. (2005). *Formative assessment: Caveat emptor*. Presented at the ETS Invitational Conference 2005, New York.
- Shepard, L. (2008). Commentary on the National Mathematics Advisory Panel recommendations on assessment, *Educational Researcher*, 37(9), 32-37.
- Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*. 78, 158-189.
- Slavin, R. (1978). Separating incentives, feedback, and evaluation: Toward a more effective classroom system. *Educational psychologist*. 13, 97-100.
- Snyder, B. (1973). *The hidden curriculum*. Cambridge, MA: MIT Press.
- Stables, K. (1992). The role of fantasy in contextualizing and resourcing design and technological activity. Loughborough University Institutional Repository.
- Stake, R. (1995) *The art of the case study*. Thousand Oaks, CA: Sage Publications.
- Stiggins, R., Griswold, M.M., Wikelund, K.R. (1989). Measuring thinking skills through classroom assessment. *Journal of Educational Measurement*, 26(3), 233-246.
- Stiggins, R. (1999). Assessment, student confidence and school success. *Phi Delta Kappan*. 72 (4): 232-238.
- Stiggins, R. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*. 83(10) 758-765.
- Stiggins, R., Arter, J. A., Chappuis, J. & Chappuis, S. (2004) *Classroom assessment for student learning: Doing it right – using it well*. Portland, OR: TES Assessment Training Institute.
- Stiggins, R. (2009). Assessment for learning in upper elementary grades. *Phi Delta Kappan*, 90(6), 419-421.

- Stiggins, R., & DuFour, R. (2009). *Maximizing the power of formative assessments*. Phi Delta Kappan, 90 (9), 640-644
- Suskie, L. (2009). *Assessing student learning: a common sense guide* (2nd ed.). San Francisco: Jossey-Bass.
- Taras, M. (2001). The use of tutor feedback and student self-assessment in summative assessment tasks: Towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26 (6): 605-614.
- Taras, M. (2002). Using assessment for learning and learning from assessment. *Assessment & Evaluation in Higher Education*, 27 (6): 501-510.
- Taras, M. (2003). To feedback or not to feedback in student self-assessment. *Assessment and Evaluation in Higher Education*, 28 (5): 549-565.
- Taras, M. (2005). Assessment – summative and formative – some theoretical reflections. *British Journal of Educational Studies*, 53 (3): 466-478.
- Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education*, 9 (2): 172-192.
- Taras, M. (2009). Summative assessment: The missing link for formative assessment. *Journal of Further and Higher Education*, 33 (1): 57-69.
- Taras, M. (2010). Back to basics: Definitions and processes of assessments. *Praxis Educativa*, 5 (2): 123-130
- Thorn, A., & Mulvenon, S. (2002) High-stakes testing: an examination of elementary counselor's views and their academic preparation to meet this challenge. *Measurement and Evaluation in Counseling and Development*, 35: 195-207.
- Thorndike, R. (2005). *Measurement and evaluation in psychology and education* (7th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Tierney, R. (2006). Changing practices: Influences on classroom assessment. *Assessment in education*, 13 (3): 239-264.
- Trotter, E. (2006). Student perceptions of continuous summative assessment. *Assessment & Evaluation in Higher Education*, 31 (5): 505-521.
- U.S. Army Command and General Staff College. (2005). *US Army Command and General Staff College Self-Study Report*. Ft. Leavenworth, KS. US Army Command and General Staff College.
- U.S. Army Command and General Staff College (2008). *Bulletin Number 40, Survey Research*. Ft. Leavenworth, KS: US Army Command and General Staff College.

- U.S Army Command and General Staff College. (2008). *US Army Command and General Staff College Intermediate Level Education (ILE) Self-Study*. Ft. Leavenworth, KS. US Army Command and General Staff College.
- U.S. Army Command and General Staff College (2010). *US Army Command and General Staff School Faculty Survey*. Ft. Leavenworth, KS. US Army Command and General Staff College.
- U.S. Army Combined Arms Center (2010) *Bulletin Number 30, The Accountable Instructional System*. Ft. Leavenworth, KS: US Army Command and General Staff College.
- U.S. Army Command and General Staff College (2011). *U.S. Army Command and General Staff School C500 Assessment Plan AY 12-01*.
- U.S. Army Training and Doctrine Command (2010) *Pamphlet 525-8-2: The United States Army Learning Concept for 2015* (2010, Sep 14, version 1.0). Ft. Monroe, VA.
- U.S. Army Command and General Staff College (2012). *Circular 350-1 US Army Command and General Staff College Catalog*, 2012 edition.
- U.S. Congress, Office of Technology Assessment (1992). *Testing in America's schools: Asking the right questions*. OTA-SET-519. Washington, DC; U.S. Government Printing Office.
- U.S. House of Representatives, Committee on Armed Services, Subcommittee on Oversight & Investigations, (April, 2010). *Another crossroads? Professional military education two decades after the goldwater-nichols act and the skeleton report*. Washington, DC.
- Valencia, S. & Anu, K. (1997). Portfolios across educational contexts: Issues for evaluation, teacher development and system validity. *Educational Assessment*. 4: 1-35.
- Vygotsky, L., (1978). *Mind in society: The development of higher psychological process*. Cambridge, MA: Harvard University Press.
- Walvoord, B. (2010). *Assessment clear and simple: A practical guide for institutions, departments, and general education* (2nd ed.). San Francisco: Jossey-Bass.
- Walvoord, B., & Anderson, V.J., (1998). *Effective grading: A tool for learning and assessment* (1st ed.). San Francisco: Jossey-Bass.
- Walvoord, B., & Anderson, V.J., (2010). *Effective grading: A tool for learning and assessment in college* (2nd ed.). San Francisco: Jossey-Bass.
- Watkins, C. (2003). *Learning: A sense-maker's guide*. London: Association of Teachers and Lecturers.
- Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment and evaluation in higher education*. 31 (3) 379-394.

- Wenger, E. (1998). *Communities of practice*. Cambridge: Cambridge University Press.
- Whetton, C., Sainsbury, M., Hopkins, S., Ashby, J., Christophers, U., Clarke, J., Heath, M., Jones, J., Punchers, J., Schagen, I., & Wilson, J. (1991). *A report on teacher assessment*. London, SEAC.
- Wiliam, D. (2000). *Integrating summative and formative functions of assessment*. First Annual Conference of the European Association for Educational Assessment. Keynote Address, Czech Republic.
- Wiliam, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*. 11 (3) 283-290.
- Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*. 13 (2), 182-208.
- Wlodkowski, R., (2008). *Enhancing adult motivation to learn: A comprehensive guide for teaching all adults* (3rd ed.). San Francisco: Jossey-Bass.
- Yin, R. (2009). *Case study research: Design and methods* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Yorke, M. (2008). *Grading student achievement in higher education: Signals and shortcomings*. New York, Routledge.

Appendix A - Interview Protocol

Understanding Faculty Experiences with Common Summative Assessments

Time of Interview: Start _____ Finish _____

Date of Interview: _____ **Location** _____

Interviewer: *Dennis S. Burket*

Person Interviewed: _____

“The purpose of my research is to understand formative uses of summative assessments by DTAC instructors. This interview will last from 45 minutes to an hour. Both of our comments will be recorded and transcribed – you will get a copy of your comments for review/comment.”

***Interviewee reads/signs Informed Consent Form.**

***Ask Survey Data questions then start recorder.**

“I want to ensure that we’re both using the same definition for summative assessment during this interview. Please look at the individual DTAC assessments in grey and mark the ones that you classify as summative.”

**(Using the DTAC Assessment Classification Worksheet w/ definitions).
Have copies of each assessment available for reference if needed.**

Questions:

- 1. Please describe how you prepare your students for upcoming summative assessments?**
- 2. Describe ways you communicate with students about the results of their summative assessments?**
- 3. What are your actions after looking at the results of your student’s summative assessments?**
- 4. What impact does summative assessment have on student motivation?**

- 5. Describe the method you use for assigning a grade to an individual summative assessment?**
- 6. What would you change about DTAC's summative assessments if you could?**
- 7. What have I not asked you about your formative use of summative assessments that you would like to add?**

Appendix B - Informed Consent

Consent Form

PROJECT TITLE: Understanding Faculty Experiences with Common Assessments of Student Learning.

APPROVAL DATE OF PROJECT: 11/26/2012

EXPIRATION DATE OF PROJECT: 11/26/2013

PRINCIPAL INVESTIGATOR: Dr. Jane Fishback, Kansas State University

CO-INVESTIGATOR(S): Dennis S. Burket

CONTACT NAME AND PHONE FOR ANY PROBLEMS/QUESTIONS: Dr Fishback, (785) 532-5554,
jfishbac@ksu.edu

IRB CHAIR CONTACT/PHONE INFORMATION:

- Rick Scheidt, Chair, Committee on Research Involving Human Subjects, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224.
- Jerry Jaax, Associate Vice President for Research Compliance and University Veterinarian, 203 Fairchild Hall, Kansas State University, Manhattan, KS 66506, (785) 532-3224.

SPONSOR OF PROJECT: Not applicable

PURPOSE OF THE RESEARCH: The purpose of this research is to understand Command and General Staff School (CGSS) faculty experiences with common assessments of student learning. This research will be used in the Co-investigator's doctoral dissertation.

PROCEDURES OR METHODS TO BE USED: I agree to allow Co-investigator Dennis Burket to ask me a series of questions about my experiences with common assessments of student learning at the CGSS. Following the completion of a brief information form, I will participate in a 30-60 minute audio recorded interview at a mutually agreed upon location. I understand that all interview materials will remain confidential.

LENGTH OF STUDY: 30-60 minute interview and review of transcripts after 60 days for transcribing.

RISKS OR DISCOMFORTS ANTICIPATED: You could be adversely affected at work if statements considered unflattering to the Command and General Staff School or your teaching department were attributed to you.

BENEFITS ANTICIPATED: Faculty concerns about Common Assessments of Student Learning are better understood and study recommendations are used to improve future CGSS curriculum. Participants will receive a copy of the completed dissertation.

EXTENT OF CONFIDENTIALITY: I understand that my answers will be treated as confidential and will only be used by the researcher for data analysis. My answers will not be attributed to me and my name will never appear in any publication. Interviews will be conducted at times, locations, and in a manner that ensure confidentiality.

IS COMPENSATION OR MEDICAL TREATMENT AVAILABLE IF INJURY OCCURS: I understand that I am not receiving any monetary compensation for participating in this research study.

TERMS OF PARTICIPATION: I understand this project is research, and that my participation is completely voluntary. I also understand that if I decide to participate in this study, I may withdraw my

consent at any time, and stop participating at any time without explanation or any adverse consequence to myself. I also acknowledge that the researcher may drop me from the study at any point.

I verify that my signature below indicates that I have read and understand this consent form, and willingly agree to participate in this study under the terms described, and that my signature acknowledges that I have received a signed and dated copy of this consent form.

Participant Name: _____

Participant Signature: _____

Date: _____

Witness to Signature: (project staff) _____

Date: _____

Appendix C - Instructor Survey Form

Instructor Survey Form

Instructor Survey for KSU Research and Dissertation Support

1. Name _____
2. Active Duty or DAC (circle)
3. Academic Rank _____
4. Years teaching in DTAC _____
5. Last AY taught in DTAC _____
6. Years as a Course Author _____
7. Years as a Lesson Author _____
8. Years as a Primary SGA _____
9. Years as an Assistant SGA _____

Appendix D - Common Assessment Classification Worksheet

Mark (x) the individual assessments that you classify as summative.
(select from the ones highlighted in yellow)

Summative assessments measure what students have learned at a chosen point in the curriculum against a standard.

C500 ASSESSMENT PLAN 13-01

C500 DTAC/DLRO C500 Exam (Individual)	
C532 Contribution to Mission Analysis	
C533 Contribution to COA Dev	
C533 COA Statement and Sketch (Individual)	
C534 Contribution to COA analysis (Individual)	

O300 ASSESSMENT PLAN 13-01

O311 Enabling/Shaping Operations Briefing (Group)	
O312 CAS/AI Quiz (Individual On-line)	
O320 Commander's Intent & PG (Individual)	
O320 Evaluate 7 th AZ Bde COA S&S (Individual)	
O320 Develop 4ID COA S&S (Group)	
O399 Exercise Performance (Individual)	

O400 ASSESSMENT PLAN 13-01

O411 Stability Operations Quiz (Individual)	
O420 LOE Assessment (Individual)	
O420 OPORD Briefing (Group)	
O499 Exercise Performance (Individual)	

Appendix E - Non-disclosure Form

Nondisclosure Statement

This nondisclosure statement pertains to peer review or transcription of semi-structured oral interviews as part of doctoral research and a dissertation by the researcher (Dennis S. Burket) and participants at the U.S. Army Command and General Staff College (USACGSC) in 2013. This qualitative research has been approved by USACGSC and Kansas State University (KSU). Each participant has been informed of the confidentiality of their participation. Any references in the dissertation will use pseudonyms to protect the confidentiality of each participant.

DENNIS S. BURKET
Doctoral Candidate
Kansas State University

Nondisclosure Agreement

I agree to maintain the confidentiality of all information and comments related to the audio recordings and/or transcripts of participant interviews conducted between the researcher, Dennis S. Burket, and the participant. I will not use or disclose any of the contents of interview materials to anyone other than Dennis S. Burket.

Signature: _____
Printed Name: _____
Date: _____

Appendix F - Kansas State University IRB Approval

KANSAS STATE
UNIVERSITY | University Research Compliance Office

TO: Sarah Jane Fishback
Dept Ed Leadership
354 Bluemont

FROM: Rick Scheidt, Chair
Committee on Research Involving Human Subjects

DATE: 11/26/2012

RE: Approval of Proposal Entitled, "Understanding CGSS Faculty Experiences with Assessments of Student Learning."

Proposal Number: 6431

The Committee on Research Involving Human Subjects has reviewed your proposal and has granted full approval. This proposal is approved for one year from the date of this correspondence, pending "continuing review."

APPROVAL DATE: 11/26/2012

EXPIRATION DATE: 11/26/2013

Several months prior to the expiration date listed, the IRB will solicit information from you for federally mandated "continuing review" of the research. Based on the review, the IRB may approve the activity for another year. If continuing IRB approval is not granted, or the IRB fails to perform the continuing review before the expiration date noted above, the project will expire and the activity involving human subjects must be terminated on that date. Consequently, it is critical that you are responsive to the IRB request for information for continuing review if you want your project to continue.

In giving its approval, the Committee has determined that:

- There is no more than minimal risk to the subjects.
 There is greater than minimal risk to the subjects.

This approval applies only to the proposal currently on file as written. Any change or modification affecting human subjects must be approved by the IRB prior to implementation. All approved proposals are subject to continuing review at least annually, which may include the examination of records connected with the project. Announced post-approval monitoring may be performed during the course of this approval period by URCO staff. Injuries, unanticipated problems or adverse events involving risk to subjects or to others must be reported immediately to the Chair of the IRB and / or the URCO.

Appendix G - CGSC IRB Approval



REPLY TO
ATTENTION OF

DEPARTMENT OF THE ARMY
U.S. ARMY COMMAND AND GENERAL STAFF COLLEGE
100 STIMSON AVENUE
FORT LEAVENWORTH, KANSAS 66027-2301

ATZL-SWA-QA

25 January 2013

MEMORANDUM FOR: Dennis Burket, KSU Doctoral Candidate

SUBJECT: Proposed Research Protocol – Understanding CGSS Faculty Experiences with Assessments of Student Learning

1. This research is being conducted to meet graduate requirements for Kansas State University. Results of the research may have benefit for CGSC. Thus, the final research document is requested for the CGSC Dean of Academics.
2. Your research has been reviewed and approved by the KSU IRB. The LD&E/CGSC HPA concurs with its findings. CAC LD&E agrees to the Kansas State University IRB as the reviewing IRB. The DoD has specific and unique requirements for non-exempt research involving human subjects. An institutional agreement must be signed by the KSU IRB
3. Your request to conduct structured interviews and/or focus group interviews of CGSS Faculty is:

Approved

Approved with Conditions (see below)

Denied (see below)

4. Should you have questions concerning the above, please contact Ms. Maria Clark, Human Protections Administrator (HPA), in the CGSC Quality Assurance Office, room 4521 Lewis & Clark.
5. You must submit a closure report upon completion of your research.

E-Signed by CLARK, MARIA L. 1005828011
VERIFY authenticity with ApproveIt
CLARK, MARIA L. 1005828011

Maria L Clark
Human Protections Administrator
100 Stimson Ave. L&C, Rm 4521
Ft Leavenworth, KS 66207-2301