An Analysis of the Relationship Between 4 Automated Writing Evaluation Software and

the Outcomes in the Writing Program Administrator's "WPA Outcomes for First Year

Composition"

by

Chadwick Le Roi Sterling

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2017 by the
Graduate Supervisory Committee:

Barbara D'Angelo, Chair
Barry Maid
Eva Brumberger

ARIZONA STATE UNIVERSITY

May 2017

ABSTRACT

My study examined Automated Writing Evaluation tools (AWE) and their role within writing instruction. This examination was framed as a comparison of 4 AWE tools and the different outcomes in the Writing Program Administrators "Outcomes Statement for First Year Composition" (the OS). I also reviewed studies that identify feedback as an effective tool within composition instruction as well as literature related to the growth of AWE and the 2 different ways that these programs are being utilized: to provide scoring and to generate feedback. My research focused on the feedback generating component of AWE and their relationship with helping students to meet the outcomes outlined in the OS. To complete this analysis, I coded the OS, using its outcomes as a reliable indicator of the perspectives of the academic community regarding First Year Composition (FYC). This coding was applied to text associated with two different kinds of feedback related AWEs. Two of the AWE used in this study facilitated human feedback using analytical properties: Writerkey and Eli Review. While the other 2 generated automated feedback: WriteLab and PEG Writing Scholar. I also reviewed instructional documents associated with each AWE and used the coding to compare the features described in each text with the different outcomes in the OS. The most frequently occurring coding from the feedback was related to Rhetorical Knowledge and other outcomes associated with revision, while the most common codes from the instructional documents were associated with feedback and collaboration. My research also revealed none of these AWE were capable of addressing certain outcomes, these were mostly related to activities outside of the actual process of composing, like the act of reading and the various writing mediums.

TABLE OF CONTENTS

LIST OF TABLES

Automated Writing Evaluation tools (AWE) are software programs that provide scoring and feedback on writing. These programs have increased in popularity as growing global connectivity and the rise of the internet increased the need for and importance of composition skills; writing skills were further emphasized when standardized tests like the SAT introduced timed writing components (Warschauer and Ware, 2006). As AWEs have become increasingly sophisticated, they offer more possibilities to the instruction of composition. Originally designed for scoring and basic feedback on writing quality, these programs are now capable of providing complex analysis of student writing by employing computer related analytics already being used in fields like "computer science, linguistics, writing research, cognitive psychology, educational data mining (EDM) and learning analytics (LA)" (Vitartas et al, 2016, p.592). The relatively recent increase in research related to these programs (Ranalli, 2017) may be reflective of the recognition by academics of the possibilities of AWE's enhanced capabilities within the practice of classroom instruction.

Previous research about AWE have concentrated on a few areas, including analysis of their overall effectiveness and suggestions about how these programs can best be incorporated into classroom environments. I could not locate any studies however, that examined aspects of the actual feedback that AWE produce or any discussions about the compatibility of this feedback and the AWE's approach towards generating it with existing pedagogical principles related to the instruction of composition. Analyzing these texts according to academic principles is one way of better understanding the relationship between AWE and the classroom environments within which they are being used. My research examines this missing knowledge within the field of AWE research and looks at

the way many of the advanced features of AWE function during a specific academic period- First Year Composition (FYC). FYC has already been recognized as a critical moment within college writing instruction by academics like Brearey (2015) and Yancey (2001), which made it an appropriate area within writing instruction to concentrate my research.

To complete this study, I used the Writing Program Administrator's influential document 'WPA Outcomes Statement for FYC' (the OS) to develop a pedagogical model for reviewing texts generated by or related to AWE as well as examine the way that these programs produce feedback. My review of this data then allowed me to identify consistencies and discrepancies between how the feedback is generated and the different outcomes within the OS. I then discuss ways in which AWE are suited to assist with meeting the goals of the OS as well as examine areas in which they are unsuitable for this purpose. The results obtained from my analysis were therefore different from previous research in that I address aspects of this relationship and deal with ideas of suitability instead of effectiveness.

My research begins with an examination of the literature related to AWEs and FYC instruction, as well as general concepts associated with the teaching composition. My review of the literature explores the accuracy of using the OS as a pedagogical model within my analysis. After which, I examine the relevant literature associated with the larger field of composition instruction; then I examine research that discusses AWE and how they are currently being used. In my discussion of my methodology, I discuss how Content Analysis was used as way to operationalize the main points in the OS and subsequently apply it to the feedback and the instructional texts associated with the

2

AWEs. Finally, I present the results of this Content Analysis and I discussed the way that this information is suggestive of larger trends between the compatibility of the selected AWE and the outcomes for FYC as outlined in the OS.

LITERATURE REVIEW

In 'Minds Online' Michelle D Miller (2014) recognizes that "instructional technology is . . .a timely concern within higher education" (p.1); Miller is expressing the widely accepted view that technology has a significant impact on education as well as the fact that this impact has also been the source of much debate. My study examined one example of this kind of technology, AWE, and its role within the teaching of FYC by concentrating on how these programs interact with the outcomes in the OS. On the surface, the scholarship surrounding AWEs and the OS do not appear to share many similarities except for the fact that they both deal with some degree of educational pedagogy. However, closer examination reveals a common trend in the arguments against as well as those in support for the use of both the OS and AWE tools.

Multiple sources have identified FYC as an important and influential educational milestone within the development of effective long-term college writing skills. In her introduction statement to the OS, Kathleen Blake Yancey (2001) recognizes that FYC "persists as a nearly universal experience at colleges and universities across the country" (p.322). In the 'Outcomes Book,' Rhodes et al (2005) conclude that these kinds of courses vary too much across institutions to be given a single definition and that the focus must instead be placed on "outcomes" or those composition related tasks that students should be able to perform after completing these courses (p.12). Finally, in the article 'Understanding the Relationship between First- and Second-Semester College Writing

Courses' Oliver Brearey (2015) makes the claim that a discussion about FYC is also necessarily a discussion about writing at all other levels. These references, as well as the fact that a segment of the academic community dedicated resources to identifying and discussing outcomes that are mostly associated with this specific aspect of education, all reinforce the importance of FYC within the overall college experience and make it an appropriate area to concentrate this research.

The WPA Outcomes Statement

The OS is "a curricular document that speaks to the common expectations, for students, of FYC programs in the United States at the beginning of the 21st century" (Harrington et al 2001, p. 323). To explain these expectations, the statement divides its outcomes into four categories:

i. Rhetorical Knowledge

ii. Critical Thinking, Reading and Composing

iii. Processes

iv. Knowledge of Conventions (2014).

These categories represent the desired outcomes agreed upon by a consensus of academic professionals. Each category subsequently expands on this list and explains what kind of knowledge students should be acquiring to become successful writers. The OS' concentration on outcomes as opposed to methods allows for the consideration of flexible pedagogical means to reach its goals; this includes approaches made possible through technological developments like the advent of AWEs. While this interpretation is not the direct intention of the Statement's lack of any specific prescriptive guidelines, it

does reflect an acknowledgement of the fact that there are multiple ways of helping

students be successful in FYC.

The Outcomes Categories.

The OS' 4 different categories represents an ideal state of knowledge for students;

they are essentially summaries of the qualities associated with mastery of a specific area.

In addition, each category is further broken down into an overview, a list of specific

outcomes that students should be able to accomplish and a guideline for how faculty in

other disciplines can help to contribute towards reaching these outcomes. By examining

each category according to the kinds of information they contain, it is possible to develop

a greater understanding of the OS and how it is related to the larger academic community

with which it is communicating. The Outcomes are also grouped into two thematically

similar sections. The first two outcomes, Rhetorical Knowledge and Critical thinking,

Reading and Composing are related to cognitive processes. These initial outcomes deal

with the idea of how to think about composition. The remaining two outcomes are

Process and Knowledge of Conventions; these are more practical and are related to how

composition actually takes place.

Rhetorical Knowledge. The OS identified Rhetorical Knowledge as the "basis

for composing." This is "the ability to analyze contexts and audiences" and then apply

this analysis towards developing appropriate composition. The role of the educator in

helping students to reach this outcome is to assist them with developing the necessary

skills for engaging in this complex mental process. It is the first outcome listed in the

statement and its physical placement coupled with the use of key words like 'basis,'

'composing,' 'genres,' and 'expectations' reinforces the idea of Rhetorical Knowledge as

5

the "uber- outcome" (Maid and D'Angelo, 2012, p. 257), considered to be foundational within the statement and subsequently within the larger discipline of teaching writing.

Critical Thinking, Reading and Composing. If Rhetorical Knowledge is the process of understanding the audience, then Critical Thinking, as defined by the OS is related to making sense of the information being presented. This outcome is defined as the ability to "is the ability to analyze, synthesize, interpret, and evaluate ideas, information, situations, and texts." It is also associated with understanding but the subject in this case is not the audience but the information available. The list of activities that characterize Critical Thinking, Reading and Composing, such as "analyze. . . synthesize. . . interpret and evaluate," are all related to the idea of scrutiny and examination.

Critical Thinking is further associated with meaning making (Dando 2016); the core characteristic of this outcome is the ability to "decode and create within a defined context" (Dando, 2016, p. 10). Mulnix (2010) however, presents a more complicated understanding of this term by stating that Critical Thinking is built around "recognizing the inferential connections that hold between statements" (p. 467). Once again the larger academic attitudes are consistent with the approach of the WPA which emphasizes outcomes related to "strategies," "question" and "evaluation." All the different perspectives share the approach of considering this outcome as being related to "how we think" (Mulnix, 2010, p.466).

Processes. This Outcome is related to the steps required to "conceptualize, develop, and finalize" the composing process. The Processes category refers to what is actually done to create a draft. This outcome is relatively straightforward and consists of "reading, drafting, reviewing, collaborating, revising, rewriting, rereading, and editing."

In this category, the OS also emphasizes other qualities such as the ability to adapt to new technologies and methods to effectively participate in the discipline within which the composition takes place.

One academic concept related to this outcome is the "Process based writing approach" which is "one of the most popular methods of teaching writing" (Graham and Sandmel, 2011, p. 396). As the name suggests, this approach to teaching composition is primarily concerned with emphasizing the mechanics of the writing activity. Although there is no single definition of this approach, it is essentially concerned with the same actions emphasized in the OS, "planning. . translating. . . and reviewing" (p. 396).

Knowledge of Conventions. Knowledge of Conventions, refers to an awareness of the genre constraints within which a document is being developed and these restrictions impact how it is created. Similar to the Processes category, this outcome is primarily concerned with decisions made while composition is taking place. The OS interprets this term very broadly and uses it to refer to "mechanics, usage, spelling, and citation practices" as well as "graphics, and document design." The Knowledge of Conventions category may further be seen as directly related to Rhetorical Knowledge, since it is the execution of the understanding developed through an awareness of the rhetorical situation within which the writing is taking place.

The most interesting thing about the WPA's definition of this category is that it is applied to both visual and written composition. This sort of application prioritizes the role of genres in developing writing and recognizes that they are defined by their 'mechanics' as much as their 'design.' In 'Changing Conventions of Writing: The Dynamics of Genres, Text Types, and Text Traditions,' Taavitsainen (2001) recognizes this all-

encompassing approach to genre by discussing how they are developed as a reaction to "external evidence in the context of culture," (p.140) and can vary significantly depending on where the communication is taking place and the format in which it is being developed. The discussion about genres and their role in determining visual and mechanical style can also be seen as a reaction to the newer electronic forms of communication which further impact composition. The WPA Statement recognizes that "genres evolve in response to changes. . . in composing technology" and Taavitsainen discusses how genres are impacted by "sociohistorical conditions" as well as "changing fashions" (p.141). These two perspectives are acknowledging the fact that electronic composition practices would subsequently have an impact on the writing process.

Literature about the Outcomes Statement

The seminal work on the OS was the 2005 publication, "The Outcomes Book: Debate and Consensus after the WPA Outcomes Statement." This collection of essays and perspectives chronicled the development of the OS from its beginnings in 1996 and its evolution over years of collaboration between Writing Program administrators and educators, to its eventual publication in 2000 (p.322). It discusses how the outcomes were agreed upon by a group of academics that represented "all levels of postsecondary education and many different kinds of institutions" (p.27); these outcomes were further refined through numerous conferences and debates. It discusses how the OS has since been used to inform the development of composition programs throughout the country; this has included a variety of applications such as defining courses, assessments, training, providing training for teaching assistants as well as adjunct faculty and even directly within the courses themselves.

Contained within the historical and critical analysis of the OS is also a discussion of the significance of the use of outcomes as opposed to standards. Both Kathleen Blake Yancey (p.18) and Mark Wiley (p.24) in their respective chapters, emphasized the flexibility of the term 'outcomes'; Yancey explains this term as "focusing on what we might call the *what* of education" (p.21) while Wiley clarifies that the outcomes are "intended to guide, not to dictate, local curricular decisions" (p. 29). In both cases, the scholarship implies that the OS is not a prescriptive document and should not be limited to any specific pedagogical or practical approach; these definitions, as well as the fact that both Wiley and Yancey dedicate their respective articles to this clarification, reflect just how important this flexibility is to interpreting and applying the OS.

In 2012, a follow up book was written that expanded and commented on the Outcomes and the original Outcomes Book, "The Outcomes Statement: A Decade later." In a chapter within this work, Debra Frank Dew's "CWPA Outcomes Statement as Heuristic for Inventing Writing about Writing Curricula" remarks that the OS retains "its historical stature as the discipline's FYC curricular statement" (p. 5); Dew's remarks highlight the continued relevance of the OS to the teaching of FYC. The article proposes a means of using the OS to update existing FYC programs to meet the new realities of academic environments which are better suited for a new kind of writing program. While Dew challenges the way that composition is taught, she continues to demonstrate how the outcomes can be extracted from the OS and applied to fix these issues. In addition to the larger curricular argument, Dew's article is further identifying the OS as being relevant to the current academic conversation.

In "The Politics of Pedagogy: The Outcomes Statement and Basic Writing" Wendy Olsen (2012) also discusses the relevance of the OS to current academic discourse but she specifically concentrates on the role that the OS can play in developing writing programs. Olsen explains how she used the different outcomes as "a curricular link that held together the sequence of writing courses" (p.25) in the then new writing program at Washington State University in Vancouver. The article further provides sample course descriptions that she developed while relying heavily on the existing verbiage of the OS. Olsen encourages embracing the "heterogeneity" (p.30) of writing and argues that the OS can provide a guide for navigating these differences and can actually strengthen the position of these programs within the academic community.

Barry M. Maid and Barbara J. D' Angelo (2012) expand on this idea of the OS as being "flexible and adaptable to other curricular contexts and types of writing programs" (p.257) through their discussion of how the OS was incorporated into the development of Arizona State University's Technical Communication Program. Maid and D'Angelo utilized the OS as a means of developing the program's curriculum, assessments, as well as a scoring guide. In this program, students were required to develop and present a portfolio as a part of their final grade; the authors examined this portfolio and coded it according to the outcomes in the OS. From this coding, Maid and D'Angelo suggest the idea of Rhetorical Knowledge as the "uber outcome" (p.257) as it seemed to be the most prevalent within their study as well as being foundational to the other outcomes within the OS.

In another relevant article from "The Outcomes Statement: A Decade Later," Sherry Rankins-Robertson suggests a process through which the OS is used to develop

student writing assignments. This article discusses how Rankins-Robertson used the OS as a guide for improving the quality of her classroom assignments and consequently better helping her students to achieve these outcomes. Rankins-Robertson's article is reflective of the many examples of the OS being used in flexible ways by different institutions. The flexibility of the outcomes has significant relevance to the larger goals of this research.

Other scholars disagreed with the emphasis that the OS, along with academics like Wiley and Yancey place on outcomes. In 'The Trouble with Outcomes: Pragmatic Inquiry and Educational Aims,' Chris Gallagher argues that these kinds of assessments "limit and compromise the educational experiences of teachers and students" (p.43). Gallagher takes issue with the fact that outcomes are generally established before the teaching process begins, as well as the fact that they may not change enough to reflect real-life developments within classrooms. However, while Gallagher's arguments may have some legitimacy, his criticism is primarily centered on how these outcomes are applied by "academic management" (p.46) and the tendency of those in power to apply these principles as a restrictive metric. These kinds of concerns are not directed towards the quality of the outcomes themselves, which are generally recognized as being positive.

The Outcomes Statement has however, also been recognized by some as "the best overview of the mainstream understanding today of what first-year writing classes should do" (Barnhisel, Stoddard and Gorman, 2012, p.462). In addition, Oliver Breary (2015) recognizes that this document is a useful way of "delineating expectations for the development of students' knowledge and skills" (p.245). Breary does express other concerns regarding the Statement's lack of specificity in some areas but like other

criticisms, there is a general acknowledgement of the OS' value. These two sources reflect the versatility with which the Statement can be applied, Barnhisel et al are making a pedagogical argument while Breary's concern is a practical one about how instruction is delivered within the first two semesters of writing specific classes. Both sources suggest the widespread application and referential use of the Statement within the academic community. In this case, both arguments supporting and criticizing the Statement reinforce its relevance.

Notably lacking from the original OS was any reference to "technology and technology's impact upon writing and literacy" (Lowe, 2006, p.225). The document did not actually address the role of technology until 2014 when it was revised to include a definition of 'composing' as "complex writing processes that are increasingly reliant on the use of digital technologies." The inclusion of this specific section is a recognition of the impact that technology has had within the field of FYC as well as a reflection of the movement within some level of academic thought towards accepting and responding to this impact. In addition, by including this definition in a pedagogical statement, it could also be inferred that there was also some acknowledgement of the role of this tool within writing instruction.

"Technology is now inextricably linked to literacy and literacy education" (Griffin and Minter, 2013 p.141); in the introduction to the OS, a similar connection is made to the related concept of composition. When the OS was revised to include references to technology, the introduction expanded the concerns of the document to include "elements of design, incorporating images and graphical elements into texts intended for screens as well as printed pages." These actions are all obviously related to

composing activities that are reliant on technology in order to be effective. A reaction to the increasing prevalence of digital technology has been a "growing number of people using technology in their classes" (Griffin and Minter, 2013, p.142). The evolution of the OS therefore, reflected a recognition of this trend and an attempt to assist with navigating these changes within the context of FYC instruction.

Feedback and Composition Instruction

There are numerous studies that reinforce the idea of feedback as an effective tool within composition instruction. This almost self-evident concept is supported by various pieces of educational research. Elham Daneshvar and Ali Rahimi (2014) conducted a study where they provided different kinds of feedback to a group of students and then instructed them to perform a writing task. The not surprising result of this experiment was that the groups of students who received feedback improved their writing to a greater degree than a control group that did not receive any feedback. However, the more interesting observation (and the point of Daneshvar and Rahimi's research) was that the kind of feedback the students received also impacted their writing. In this case, feedback which encouraged the students to review then rewrite and entailed "guided learning and problem solving. . ." as well as "reflection" (p.218) was more effective at improving student writing than more generic feedback that focused on form and structure.

In 'Effects of the Red Pen,' Harriett D. Semke (1984) discussed the role of feedback in writing instruction and while the study concentrated on foreign language learners, Semke also drew some universal conclusions about general composition instruction. Semke addressed the "time consuming" (p.195) nature of providing feedback as well as the frustrations of teachers who worry that their efforts are going unnoticed or

are ineffective. Semke actually arrived at a surprising conclusion which is that "'correction does not appear to promote competency" (p. 202) and that the time teachers spend fixing student work is essentially wasted. Semke advocated that the more effective means of developing student writing skills is through facilitating practice. The article does however, also recognize that there is some benefit to be gained from providing feedback that takes the form of comments.

While these two arguments may at first appear to be contradictory, there is a common theme between the research done by Daneshvar and Rahimi and the ideas advocated by Semke. Daneshvar and Rahimi's research actually supported the kind of feedback that was advocated by Semke; feedback that encourages rewriting. Both articles agree that there is some value in teacher's providing feedback regarding student writing but the difference is that Semke makes an analysis of whether the benefits that students get from feedback is enough to justify the effort and time teachers exert to provide feedback to every student.

The overall importance of feedback to composition instruction is further evident in the educational concept of 'Process-based writing pedagogy.' Process-based pedagogy is built around the idea of "recursivity" (Barnhisel, Stoddard and Gorman, 2012, p.463), which involves the repeated revision of writing in order to develop a clear idea. This revision-centric process is based on "student response to formative feedback" (p.464). Process based pedagogy "has for three decades been the standard model used in postsecondary writing programs" (p.463). It encourages students to respond to feedback on their writing while still composing and then review additional feedback from their professor once the composition process has ended; this final feedback is also generally

associated with a grade. The prevalence of this method is perhaps the clearest indicator that feedback has already been accepted as a suitable means of achieving outcomes within the academic community.

George Hillocks Jr. (2007) developed another theory of instruction that was specifically related to narrative writing but can also be applied to the more general practice of teaching composition. Hillocks addressed the idea of feedback directly as well as indirectly by suggesting tools that could be developed to assist students and teachers navigate the composition process. He identified the ""Environmental Mode" of writing instruction" which emphasized "student choice supported by specific objectives, structured collaboration. . . and activities selected by the teacher" (Gorlewski, Krickovich and Gorlweski 2011, p.110). This approach to writing instruction is unique because it is both teacher centric while also being concerned with the role of students in improving their own writing. Hillocks presents three criteria for assessing composition: "specificity, style, and (episodic) elaboration" (p.110); he also encourages the use of rubrics at every stage of the writing as a way of helping students engage in revision and improve their composition. He suggests a checklist to assist students as they are participating in revision. In addition, Hillocks engages with the idea of what kind of feedback is most effective and appears to align more with Semke regarding the importance of positive comments and the ineffectiveness of too much detail.

AWEs can be seen as having been developed in response to some of the concerns that were voiced by Semke as well as based on the approach to feedback that was advocated by scholars like Daneshvar and Rahimi while also containing some of the central elements associated with Process-based learning. Essentially, these kinds of

software are, among other things, an attempt to reduce some of the pressures associated with providing detailed feedback on student work. AWEs actually provide the two things that Semke identify as being important for learning and that are also core aspects of Process-based writing, these are an opportunity to practice and the chance for feedback. The emphasis that these software place on feedback within the learning process is obviously more significant than that which is employed by Semke and is more consistent with Daneshvar and Rahimi's perspective.

Automated Writing Evaluation Tools

AWE are software that provide "computer-generated feedback on the quality of written texts" (Stevenson & Pakhti, 2014, p.52). The birth of these programs, like other movements related to the teaching of writing, was a result of the "broader social, economic, cultural and political contexts that shape the needs for using the language" (Warschauer and Ware, 2006, p.157) Most AWE are web-based and allow students to create an account and interact with content developed by their instructor (Ware, 2011, p.770); these programs "employ natural-language processing, machine-learning or other computational methods in the analysis of text" (Ranalli, Link and Chukharev-Hudilainen, 2016 p.8). Warschauer and Ware discuss the birth of AWEs in 'Automated writing evaluation: defining the classroom research agenda'; They trace the origins of AWEs back to the 1960s when a group of universities developed Project Essay Grade, a program designed to score high school student writing (p.158). Project Essay Grade however, was limited by the technology available at the time. Years later, the software was updated; a newer version was released along with the Writer's Workbench, another program that provided automated writing feedback based on a limited set of criteria.

There have been significant technological improvements since the early stages of AWE development. "Innovative analytical tools have allowed teachers and educational designers to understand student performance in much greater detail" (Vitartas et al, 2016, p.592). These developments have subsequently complicated the way that these kinds of software are viewed as well as their perceived usefulness within classroom environments. AWEs are now capable of performing sophisticated tasks like "discourse coherence," "source use and integration," and "topicality" (Burstein, Elliott and Molloy, 2016, p.118) which has led to new conversations regarding exactly how these new capabilities should be used. This conversation about how and even more fundamentally, whether or not to use these programs has dominated the field of AWE literature. Such debate has subsequently stratified based on how the software is attempting to apply itself to educational environments.

In response to the growing complexities of AWEs, Burstein, Elliott and Malloy (2016) suggested a role for this software that was more involved than simple analysis of academic essays. They stated that AWEs should actually "focus on social and rhetorical knowledge, domain knowledge and conceptual strategies, the writing process, and language use and conventions." (p.135). The similarities between what Burstein et al suggest as the preferred role of AWEs and the Outcomes advocated by the WPA are obvious. While their survey did identify the OS as one source in its "continuum of consensus opinions on writing competencies in school" (p.128), it also further developed this perceived role for AWEs by examining qualitative data which identified shortcomings in the current practice of composition instruction.

Despite the significant technological improvements that have been made since the early stages of Project Essay Grade and the Writer's Workbench, the two uses for which these AWE were developed have remained the primary way that all future programs have been applied. AWEs still tend to fall into two categories related to their purpose; they are either focused on assessment or with providing feedback. Scholarship about these kinds of programs is subsequently divided into similar categories; even as Academics have recognized the value of automated feedback, there is still uncertainty and debate as they attempt to identify a place for this tool within writing instruction. While there is much conversation about the more controversial scoring component of AWEs, very little scholarship exists about the benefits of their feedback generating ability (Stevenson, 2016). Any information available is further impacted by the fact that most research in this area is "largely funded by the companies that have produced commercial AWE products and carried out by staff of these companies" (Warschauer and Ware 2006, p.163). This issue of credibility related to AWE research was also expression by Stevenson (2016), who worried that most research conducted to determine the validity of these programs was conducted by "researchers affiliated in some way with companies that develop and market AWE systems" (p.2). Examination of the available data shows that "while the scoring systems' validity remains contended, their diagnostic feedback function seems pedagogically appealing for formative learning" (Chen and Cheng, 2008 p.97).

The debate about AWEs is in many ways similar to the conversations regarding the OS and reflect a general belief that objective or standard analysis is not useful when examining at the nuances of writing. Vitartas et al (2016), suggested that a move towards more assessed writing is something that AWE developers should consider as positive for

their field (p.598); such a move would undoubtedly mean that the kind of analysis offered by these programs would become more necessary. When Kathleen Blake Yancey and Mark Wiley discussed outcomes and standards, each warning against the dangers of one or the other, they were reflecting the same kind of distrust that characterizes much of the critique against AWE tools. Yancey (2000) states that standards stifle creativity and that meeting a standard does not necessarily mean that that the student had attained proficiency in the manner intended. This is similar, in some degree to a worry that will be discussed later, which is that students could possibly 'trick' an AWE by developing a polished nonsensical draft. The WPA developed Outcomes instead of standards a way to avoid this issue (although Wiley argues against this choice as well), similarly, AWEs take steps to try and adjust for this potential weakness. The scoring systems and the feedback generating software have different criticisms and they attempt to compensate for them in different ways.

AWEs as Scoring Systems

While the scoring component of AWEs is not the focus of this research, it is still important to understand the debate surrounding these programs, because some of the concerns as well as the arguments in the support of these scoring systems can also be applied to the feedback generating software that will be examined in depth. The scoring features of AWE offer immediate advantages for assigning grades to papers, these advantages are primarily related to expediency and savings associated with "money, time, objectivity, and reliability levels" (El Ebyary and Windeattt, 2010, p.123). The concern therefore is whether the efficiency of these products is also reflective of their effectiveness. The bulk of research in this area has concentrated on the correlation

between the scores provided by the AWEs and those given by human raters; when this measurement is applied, these kinds of software generally perform favorably in relation to their human counterparts.

According to Warschauer and Ware, while the bulk of existing research on AWEs is not necessarily objective, generally a computer-generated score on a standardized test will be very close (within 1 point) to that of a score given by a human rater at least 95% of the time. However, these results are expected to be less favorable when the scoring is done within classroom environments where the content of the writing is more important. There has, however, been a notable exception to the standardized testing rule, the IntelliMetric system assigned grades that were significantly higher than those given by human rates; human raters gave failing scores to 27.1% of students while the system only failed 2.8%. (Vitarats, 2016, p.597) In addition, Warschauer and Ware also determined that the relationship between computer generated scores and other factors like GPA in writing courses was not as strong as the relationship with scores assigned by human raters. These complex discrepancies related to the effectiveness of these programs seem to reinforce the concern of academics like Ranalli, Link and Chukharey (2016) that the scoring component of these software cannot capture the complexity of language.

Despite the many issues identified with these kinds of AWE, they are relatively widely used. Criterion, specifically, is used to grade the Test of English as a Foreign Language (TOEFL) as well as the essay portion of the Graduate Management Assessment test (GMAT). The software however, is used in conjunction with a human rater and any discrepancies between the two are automatically sent to another human rater to be resolved (Warschauer and Ware, 2006, p.161). This kind of official use is

surprising given the general skepticism suggested by bulk of scholarship related to these programs but in this case, their advantages seem to encourage academics to try and adjust for whatever shortcomings may exist.

AWE Feedback Generating Components

The focus of my research is the feedback generating component of AWE software. Scholarship in this area is more limited but opinions towards these kinds of software tend to be mixed and are in some cases, more favorable than those towards the scoring component of AWE. One of the reasons for the acceptance of this category of AWEs by the academic community appears to be related to the fact that the bulk of these programs are designed to supplement and not replace traditional classroom instruction, which removes many of the criticisms associated with feedback generated solely through computer analysis. A common characteristic of all feedback generating AWE is that "they provide writers with multiple drafting opportunities" (Stevenson and Phakti 2014, p. 52). Another reason for the relatively positive attitude toward these programs is that they reflect a pedagogical preference; these kinds of software support a tendency to question the "effectiveness of today's standardized testing as the basis for educational accountability" while emphasizing "timely and actionable feedback about student learning" (Cope et al, 2011, p.80). The differences between these AWE can generally be found in the type of feedback that they provide and the way this feedback is delivered.

Feedback generating AWE are, however, also subject to many of the same criticisms as their scoring-centric counterparts. This idea that computers are not properly equipped to evaluate writing in a complicated manner would also obviously be a concern when considering whether the feedback given by these programs can be trusted or

considered effective. However, also like their counterparts, these programs offer benefits

related to expediency and therefore have also been carefully allowed into the classroom,

generally serving as a resource for instructors. Feedback generating software have also

been identified as capable of helping to "improve learning outcomes" (Cope et al 2011, p.

85), primarily because they can provide immediate and constant feedback in a manner

that human instructors are simply incapable of doing on a large scale. In addition, these

programs also allow for instructional flexibility by providing an option for individualized

learning to large student groups who are not always performing at the same level.

 Ware (2011) supports this view of AWE tools as supplemental resources for

writing instructors. However, Ware also points out that while using AWE tools coupled

with classroom instruction have been shown to increase test scores, academics are still

concerned that employing a computer-based method for assessing writing will result in

composition that is "mechanistic and formulaic" (p. 771). Ware's research determines

that there were observable benefits to student writing if they interacted with these sorts of

software consistently and for a prolonged period. Like others, Ware determined that this

kind of feedback encouraged improvement in a specific area of writing which tended to

be "mechanistic and formulaic" (p.771). While the OS included Knowledge of

Conventions as one of its categories, Ware remained uncertain about whether the benefits

of these programs were equivalent to the opportunities for instruction teachers would give

up when they allowed students to engage extensively with AWEs during time that could

be used otherwise. Ultimately, the article chooses to resolve this ambiguity about the

effectiveness of AWEs by suggesting instead that each academic institution determine

how these programs could be aligned with their own values and concentration as opposed to applying any kind of universal standard.

Other case studies look at AWE tools and the benefits of their feedback generating components for helping students develop their writing abilities. Chandrasegran, Ellis and Poedjosoedarmo (2005) examined one specific program, Essay Assist (p.137) within a classroom environment. This examination attempted to determine if Essay Assist could help students with their writing in a more sophisticated manner; the feedback given was designed to help the writers consider the 'socio-rhetorical situation of a writing task' (Chandrasegran et al 2005 p.147) and then use this perspective to increase the quality of their composition. While this study is dated, it is relevant because Chandrasegran chose to define 'socio-rhetorical' as related to participation in the "on-going conversation of the disciplinary discourse community" (p.139); such a definition is very similar to the OS's description of Rhetorical Knowledge. Chandrasegran's research may therefore be seen as an indirect and likely unintended case study which examined the way that an AWE software attempted to help student reach one of the WPAs outcomes.

Essay Assist was designed to provide students with feedback on their writing and essentially guide them towards the kind of thinking necessary to develop effective essays. The study distributed this software to a group of students and then drew some conclusions about their experiences with the program. This research had several shortcomings however, including the fact that the participants were not monitored during the interaction with the software which meant that the findings were almost entirely based on subjective accounts. In addition, the authors acknowledge that 'technical problems' (p. 147) impacted the students experience and would have undoubtedly

influenced the final evaluation of the program. However, criticism related to the usefulness of the feedback was more mixed and many of the students did feel as if the software was giving them valuable feedback but this experience seemed to have been impacted by the program's usability issues. The main takeaway from the article was that it is possible, according to the students who participated in the study, for software to provide feedback that has a positive impact on guiding students towards developing rhetorical knowledge. It is also important to consider that technology in this area has evolved a great deal and it is possible for these programs to operate in a much more effective manner than at the time of the original study.

Another, more recent case study attempted to examine general perspectives within the academic community towards writing and then determine what role AWEs could play in meeting the current needs. Burstein, Elliot and Molloy (2016) conducted 2 surveys and lead a discussion group during which they gathered the views of educators, students and workplace professionals regarding what each group considered to be most important about how writing was being taught and subsequently practiced. These results were then used to make observations about the current state of writing instruction as well as to develop suggestions about ways in which AWE can contribute to how writing is taught. The research found that there was a discrepancy between the 'transactional genres' taught in high schools as teachers try to prepare students for the workplace and the kinds of 'expressive' genres they have to learn in order to actually perform the complex tasks required to be successful in a post-secondary environment (Burstein et al, 2016 p.133). Ironically, the criticism of the existing academic system is actually consistent with the concerns that Ware (2011) expressed regarding the kind of writing that would be

encouraged if students were to receive their grading and feedback from a computer program. If Burstein et al's criticism is to be accepted, then human instruction can also lead to writing that was described by Ware (2011) "mechanistic and formulaic"; this simply means that the format in which the instruction is delivered may not necessarily be the determining factor for effectiveness in composition instruction.

Burstein et al's research obviously took place at a time when the feedback generating capabilities of AWEs were much more sophisticated. Their research identified a shortcoming in the existing academic instruction and determined that "students were lacking significant command of core competencies" (p.132). This criticism however, was directed at the way that instruction was being delivered within classroom settings and identified as a way AWEs could respond to the needs of college professors and even workplace professionals who felt as if students were not being prepared to perform the kind of writing considered valuable in their respective areas. The main premise of Bustein et al's argument was that as AWEs become more sophisticated, they can fill gaps in the existing educational system. The article recognized the capability of this software to accomplish these goals but does not necessarily elaborate on how this should be done.

The increasingly sophisticated nature of AWE software and the potential of this change is further discussed by Noreen S Moore and Charles A. MacArthur (2016) in "Student Use of Automated Essay Evaluation Technology during Revision." Moore and MacArthur discuss how these technological advances now offer greater opportunities for teachers to evaluate student writing and encourage increased feedback. While Moore and MacArthur emphasize the benefits associated with expediency just like previous scholars, they also discuss the increased benefits associated with their advanced capabilities;

25

Moore and MacArthur argue that these programs can use their combined scoring and feedback generating capabilities to encourage students to participate in revision. Similar to other scholars, Moore and MacArthur also suggest that these programs should be deployed in conjunction with traditional classroom instruction as a way to overcome any possible issues with the way that they work.

Vitartas et al (2016) adopt an equally optimistic tone about the role that AWEs can play in providing feedback and actually helping to accomplish sophisticated functions with regards to providing feedback and even scoring essays. Vitartas et al examine the tools that developers of these programs now have at their disposal as they attempt to analyze and respond to text in a meaningful manner. They discuss phenomena like "Learning analytics," "Natural Language Processing," and "Latent Semantic Analysis" (p. 593-4). These terms all refer to machine processes that make it possible for AWE to analyze student writing and respond in a manner that was not previously possible. These tools are primarily geared toward the kinds of analytical processes that can only be accomplished by a machine. This article suggests that the value of this data is enough that these educators should pay closer attention to the possibilities of incorporating AWEs into their classrooms and using these tools to help improve their ability to teach.

In the next chapter, I discuss the methods for my research and explain how I utilize this knowledge about composition instruction and AWE in order to develop an appropriate coding process to analyze these software and the feedback that they generate.

METHODOLOGY

My review of the existing literature supported the idea that the OS represents accepted pedagogical perspectives within the academic community and is therefore an appropriate document to use as a means of understanding the role of AWEs in FYC and subsequently as a basis for coding. I therefore, concentrated my research on the kind of feedback encouraged/ generated by AWEs and whether this feedback was consistent with the outcomes in the OS. I selected 4 AWEs for this study, each of which represented one of two common ways that these software are currently being used to provide feedback on student writing; either by directly generating automated feedback or by enhancing the ability of teachers and peers to provide their own feedback. I employed a result oriented approach to this analysis of AWE tools and used these conclusions to make assumptions about what role this software could play in meeting the outcomes outlined by the OS.

My research was constructed around the use of the OS to develop a single objective standard of coding which was then used to evaluate the selected software. To increase the representative quality of my research, I concentrated on the kind of texts generated by and related to these software; I did not address or consider subordinate issues related to technical glitches or user familiarity with the software since these considerations were beyond the scope of my study.

Content Analysis

I employed Content Analysis as way to analyze the relationship between AWEs and the OS. My analysis concentrated on the sample feedback generated by these programs as well as instructional materials retrieved from each corresponding software's website. Content analysis allowed me to directly compare the software's performance

against the outcomes in the OS. I used the OS to adopt a deductive approach in my analysis, using the outcomes to organize my interpretation of the data.

The Outcomes Statement

I selected the OS as the basis of coding, primarily because my review of literature associated with this document supported the idea of flexibility of application as being a key part of its intended use. While I also encountered some criticism directed towards this document, this information was not sufficient to undermine its overall credibility. The OS was an effective reference, in part because of the debate and scrutiny to which its outcomes had been subjected. These outcomes effectively represent the evolution of academic thought regarding the teaching of FYC, having been established through consensus by professionals in this field.

In conducting the Content Analysis, I operationalized the outcomes in the OS and used them to develop the codes. This was effective because the OS already included categories and information in a manner that allowed for a readily available rigorous coding format. I used the different categories in the OS and their corresponding bulleted points to develop an already objectively proven analytical construct and then used it to code the feedback generated by the AWE. Once this process was complete, I could then draw conclusions about the relationship between the coding and the selected texts.

Data Collection

To complete this research, I requested writing samples from the Arizona State University Director of Writing Programs on the Tempe campus.

Student Sampling. My research was therefore based on writing samples from 10 anonymous students who had participated in ASU Tempe's First Year Writing Program. Students in the First Year Writing program submit final portfolios to an electronic database hosted through Digication; as part of this process, students agree to submit to a digital repository and sign informed consent allowing their work to be used for research purposes.

Document Sampling. The 10 samples of student composition were all narrative pieces, approximately 2 pages in length. When I received these documents, the ASU First Year Writing program had already assigned two-letter codes to each of them as a means of identification. The 10 documents used in my research were identified as: KB, YZ, AM, AW, DC, BT, MH, RM, JD and DM.

Coding Methods & Analysis

In addition to using the OS to develop a codebook, I also used QDA Miner, a qualitative data analysis software to assist with examining the qualitative and quantitative characteristics of the final data. This program allowed me to upload multiple documents as separate files that could all be coded using the same codebook. In addition, it assisted me with easily and quickly applying codes to each line of every document. Finally, QDA Miner allowed me to quickly compare the data using graphs, charts and lists in a way that was useful for the analysis of all the different types of information.

Developing the Codebook. I developed the codebook using the outcomes in the OS. I used an a priori, theory driven approach to simplify and reword the definitions associated with each outcome in order to make them appropriate for describing the way that the selected pieces of feedback were encouraging students to interact with their

writing. I developed the codebook in a manner consistent with a process outlined by James Thomas and Angela Harden (2008). These scholars described the 3 stages for developing a codebook built around identifying thematic similarities in "Methods for the thematic synthesis of qualitative research in systematic reviews":

Stage 1: The coding of text line by line

Stage 2: Developing descriptive themes

Stage 3: Generalizing analytical themes

The OS provided me with the structural and conceptual information necessary to impose an a priori framework on the coding process. Using this information, I produced 19 codes developed from the outcomes immediately following each occurrence of the phrase "By the end of first-year composition, students should." I subsequently labelled the codes according to the category with which the corresponding bulleted point was associated. I labelled all the codes related to Rhetorical Knowledge as RK (1 thru 5), those related to Critical Thinking, Reading, and Composing were coded as CT (1 thru 7), those dealing with Process were labeled P (1 thru 7) and finally I labelled the Knowledge of Conventions section as KC (1 thru 6). Once I established this framework, I applied these codes to each line of the selected feedback.

Through this coding process, I further reworded and restructured the definition of each code to make it appropriate for use as a descriptive means of assessing text. The decisions I made during the development of these descriptions meant that some texts had to eventually be recoded to establish consistency and that I was unable to apply those codes whose definitions were not appropriate for my purpose.

Finally, my new definitions for these codes remained thematically similar in a manner that was consistent with the existing categories of the OS. Therefore, I was able to continue using the existing categories of the OS to group these codes together. Table 1 is the Codebook developed at the end of this process, it includes the 19 codes and their corresponding definitions.

Table 1 Completed Codebook

Processes	
RK 1	Any comment that encouraged changes to writing related to the rhetorical conditions of logos, pathos and ethos. Comments related to clarity and feedback geared towards improving the strength of arguments (logos) as well as comments about how to make the writing more exciting or colorful (ethos).
RK 2	Comment about using genre to fit the rhetorical situation. In the instructional material this was also statements about how the software could help teachers to update assignments since this meant that they could change the rhetorical situation that the students were responding to.
RK 3	Comments related to how the sentence could be read or the 'reader' or 'audience.' Comments encouraging additional approaches to writing and discussing how word changes could impact the meaning of sentences. Passive/ Active Voice changes were included as well as simplifying sentences options. Changes to word choice or structure that were suggestions and not simply a command (for example, delete: word).
RK 4	Any reference to the medium in which the writing is taking place
RK 5	Comments about what makes a medium (print/ electronic) appropriate for a situation
Critical Thinking	
CT 1	Comments about using writing to help develop understanding/ encourage thinking about process of writing
CT 2	Comments that encourage reading a variety of texts to increase understanding
CT 3	References to finding and evaluating research. Comments about dealing with evidence, sources, developing arguments.
CT 4	Comments on summarizing, synthesizing; statements considering how the student interpreted an idea.
CT 5	Comments that support thinking about the discipline within which the writing is taking place.
CT 6	Comments about evidence and making sure it is appropriate for discipline/ subject.

CT 7	Comments that encourage reading discipline specific text and information.
Processes	
P1	Comments about doing several drafts. Like RK 3, comments that discussed how or encouraged the student to try different ways of composing the sentence
P2	Comments about revising, collaboration etc.
P3	Comments as a means of reconsidering and rethinking ideas (meaning)
P4	Comments that discuss collaboration/ encourage multiple writers to interact
P5	Comments that discuss and encourage feedback; statements that explain how they should be used
P6	Comments about learning and using different technology such as MS Word, blogs, Adobe products, etc
P7	Comments about how different composition practices impact the work
Knowledge of Conventions	
KC1	Comments about fixing grammar issues. Passive/ Active etc.
KC2	Comments about paragraph structure (e.g. no topic sentence) or about organization (e.g. there isn't an introduction or a conclusion or paragraphs aren't arranged logically)
KC3	Comments about conventions in different genres (likely not possible using the software)
KC4	Comments about design/ formats
KC5	Comments that discuss use citations
KC6	Comments that refer to the conventions of citation.

AWE Software

My research concentrated on 4 AWEs. I selected these programs based on three

criteria: their ability to be immediately accessed and a relative degree of mainstream

representation, generally implied by the fact that they were already being used by

multiple academic institutions. In addition, these programs also represented the two

different ways that the automated functions of AWEs were being applied within

classroom environments.

Two of these programs used automated functions like computer-based analytics to

enhance the human feedback process:

- WriterKey

- Eli Review

While the other 2 programs generated automated feedback:

- PEG Writing Scholar

- WriteLab

Selecting these two different types of feedback generating AWE further enhanced my analysis by allowing me to collect data about and subsequently discuss the relationship between the various applications of these kinds of AWE. In order to gather the necessary information, I applied the codes to the specific feedback generated by WriteLab and PEG Writing Scholar, as well as to the instructional information found on the websites of all 4 programs. This allowed me to draw comparisons between the way the individual programs interacted with the categories in the OS. Below is an overview of the different programs and the way that I collected the text for coding:

WriterKey. WriterKey is perhaps one of the most obvious examples of the intersection between AWEs and classroom instruction. It emphasizes accommodating human feedback instead of providing an alternative. WriterKey allows students to upload their work to the program and receive direct feedback from their teachers and peers. Students can then engage with this feedback as they work towards completing their writing project. Writerkey's main component is its ability to review the contents of a student's essay and then use its analytical functions to assist with providing feedback.

Since WriterKey's features were not consistent with the kind of analysis performed by WriteLab and PEG Writing Scholar, I concentrated my analysis of this

33

software on a representative instructional document from its website. I decided to select text from the website instead of information found elsewhere because this kind of information was intended to represent those functions that the software claims to be capable of performing. I used the information from the "WriterKey Overview" webpage because it outlined the principles behind how the program worked.

Eli Review. Eli Review allows teachers to develop small writing tasks for students as they work towards a final composition goal. Students develop their writing with the help of feedback from their peers and guidance from their teachers. Instructors can view this information in the form of analytical data which is then used to help them develop more effective lesson plans. Eli Review's literature emphasizes the role of its analytical functions in improving the overall process of teaching writing across various disciplines. The website claims that this program allows instructors to review how useful students have rated the software it already received as well as track the level of students engagement with the writing process. This emphasis on data, coupled with its unique style of developing large writing projects in small stages means that the pedagogical perspectives and other features emphasized by this program will likely be different from those associated with WriterKey.

Like WriterKey however, Eli Review did not actually generate any feedback to analyze. I concentrated the analysis on a similar piece of instructional text that was retrieved from the website. In this case, I selected the information under the heading "How Eli Review works;" this text was also an overview of the principles behind Eli Review's design and how it was intended to improve the student and teacher feedback process and subsequently the quality of student writing.

PEG Writing Scholar. PEG Writing Scholar allows users to select from three different kinds of essays, argumentative, information/ explanatory and narrative; once an essay type is selected, then students can begin writing. This program also includes a pre-writing section where students can outline their main ideas, prior to beginning the writing process. Automated feedback is not given about the pre-writing section but a writer can receive as many comments as they would like on their draft as PEG also includes a 'revise' option. PEG generates two kinds of feedback; the first deals specifically with the text (Spelling and grammar issues) while the second is a report which assigns a score to student writing along with general suggestions for improvement.

I obtained a free trial of PEG by contacting their sales department. PEG was designed to work with instructors as a part of their class and so to facilitate my analysis, the PEG representative created a standalone class which then allowed me to access the program and upload the sample documents. I did not use the pre-writing section of the software because it was not relevant to my research. The program allowed students to respond to specific instructor created writing prompts so to generate feedback I needed to access one of the generic 'Student Choice' options. To get appropriate feedback from PEG, I also needed to specify what kind of essay was being uploaded; in this case, all of the essays belonged to the narrative category. The resulting feedback took the form of highlighted perceived spelling and grammar issues as well as a 'score report' that assigned a numeric value to the essay. This report also included a graph that represented 6 different 'traits': Development of ideas, organization, style, word choice, sentence structure and conventions; finally, text based feedback related to the 6 different traits was also included.

I located a representative instructional document on the PEG Writing Scholar website under the section labelled 'About PEG Writing Scholar.' Like the other software documents, the information contained in this section also outlined the way that PEG was intended to be used by the company, emphasizing those aspects of the programs considered to be most useful.

WriteLab. WriteLab concentrates solely on generating and facilitating feedback; it does not include a scoring component. This software is advertised as being developed in collaboration with students and teachers; it emphasizes the fact that the feedback it provides is more sophisticated and nuanced than simple surface level corrections. It claims to use the Socratic method to encourage better composition through questions geared towards identifying and creating improvements in 4 different areas: Clarity, Logic, Concision and Grammar. In addition, this program is also advertised as a resource for teachers since it's automated feedback process can also be complemented by peer and instructor comments. The two programs differ in that WriteLab is not limited in the prompts to which students can respond and its feedback emphasizes the importance of elements related to style.

I also had to request a free trial to access WriteLab. In addition, WriteLab contained an essay planning feature which I also chose not to use. To generate the feedback, I uploaded the sample essays one at a time. WriteLab then generated detailed feedback according to the 4 different areas that it had identified as important to improving student writing. Similar to the other programs, I also retrieved an instructional document discussing the ideas behind how WriteLab was intended to work. This information was retrieved from a section of the website labeled "How it Works."

Coding Process

  The process for retrieving and coding the feedback and the instructional documents are outlined below:

 1. Before I could apply the codes, I individually uploaded the 10 samples of student writing into WriteLab and PEG, the two programs capable of providing software-generated feedback.

 2. Once I retrieved this feedback, I separated it from each of the writing samples by copying the text of the feedback into individual word documents.

 3. I labelled these 10 standalone documents according to the sample text from which it had been generated. Therefore the WriteLab Feedback related to sample text AW was subsequently labelled WriteLab AW etc.

 4. I included all the WriteLab feedback in these documents.

 5. For PEG, I included all the feedback it generated directly over the text as well as all the text from the associated Score Report that had been labelled as 'Feedback.'

  The process for coding is outlined below:

 6. I subsequently uploaded these documents to QDA Miner, grouping the instructional texts into a single file and then separating the WriteLab and PEG feedback into two files as well.

 7. I concentrated the coding on each unit of feedback, however, I did not limit any piece of feedback to a single code; one piece of feedback could be coded several times in many ways because these codes were not mutually exclusive.

 8. The coding was applied to every piece of feedback in the WriteLab and PEG Documents.

9. A similar process was used to code the instructional documents, however in this case I limited the unit of analysis to a single sentence.

While I was applying the codebook, it also became apparent that some codes could not be applied to feedback from the software. This was because some of these outcomes dealt with the texts in a manner that was inconsistent with the individual, single document approach adopted by this study. These observations were still valuable to the purpose of this research; this phenomenon was discussed further in the Results and Discussion sections. This was also something that I observed when examining the instructional documents although there were some minor variations.

Limitations

The most obvious limitation of my research was my inability to analyze the kind of feedback that was facilitated by WriterKey and Eli Review. Despite the similarities between the instructional documents for Writelab and PEG Scholar and the actual software, my discussion about how these kinds of AWE operate will obviously be incapable of effectively analyzing the nuances associated with exactly how these programs help to create feedback. This means that while I can generally discuss the way the 2 programs that accommodate human feedback interact with the AWE, a more specific analysis would require a case study or some other research method that includes observations of how people interact with these programs.

Another possible limitation of this study could be related to the fact that my examination of feedback and instructional documents focuses on these documents as standalone texts; I do not consider whether these documents are accurate or appropriate. My primary purpose in evaluating these texts was to determine if these AWE can play a

role in helping students meet the outcomes in the OS; I did not consider their effectiveness in accomplishing this purpose which would require a more evaluative kind of research and analysis.

However, even research developed for evaluative purposes by employing case studies to collect data and included a wide range of participants were still capable of being impacted by outside factors like whether students choose to use the software, the kinds of technical issues they experienced while using these programs and other conditions which were not useful for the kinds of information that I was attempting to determine. By concentrating entirely on the WPA OS as an academic text and on the feedback from the AWE as the practical examples of how these programs work, I was able to get useful, result oriented data that is based entirely on objective information. While there is always a degree of subjectivity associated with Content analysis, a recognized advantage of this method is that it allowed me to "reveal trends and themes" even while limiting my ability to determine "cause" (Wilson 2011 p.43). Since the goal of my research was to determine if the kinds of feedback provided by AWE were consistent with the outcomes in the OS, the strengths of my approach outweighed any possible weaknesses.

I extracted both qualitative and quantitative data from the feedback-related and instructional texts after refining and applying the codebook. The collected data allowed me to compare the two different kinds of AWE and the way that each of them interacted with the OS. This information provided insight into the pedagogical preferences that informed the way these programs are developed and further comparisons of these preferences helped me to quantify the nature of the relationship with the OS. In the next

section, I outlined the results of this comparison and then discussed the nature of this relationship by examining how these different pieces of information interact with each other.

RESULTS

My research was primarily concerned with the nature of the relationship between the OS, represented by the codes, and the texts to which I applied this coding. My analysis of the results therefore concentrated on the similarities and differences between these pieces of data as well as the consistency and frequency of the codes identified throughout the different texts. This approach allowed me to compare the different AWE based on how they responded to the coding as well as make a larger comparison between the AWE feedback and instructional documents and the outcomes within the OS. Once I established this process of identifying and representing these relationships, I was subsequently able to discuss their significance.

My research expands on earlier work such as the 2016 study by Burstein, Elliott and Malloy that used existing academic concepts as a means of suggesting a role for AWE within classroom environments. I also evaluate these tools, similar to Chandrasegran et al's (2005) study of Essay Assist and other case studies that analyze the effectiveness of specific AWE. However, I do not perform an actual case study as a part of this research and instead I employ a theoretical analysis to draw larger conclusions about the way these programs operate; in this way, my approach is also similar to Ware's (2011) discussion of how these programs should be used by professors. Ultimately, my study builds on these earlier discussions about AWEs; I am primarily concerned with the compatibility of these tools with traditional teaching methods. The larger goal of my research is to advance this

conversation through analysis of the relationship between the selected 4 AWE and the OS; in doing this, I also provide a framework for understanding this relationship.

This analysis provided me with some insight into the relationship between AWEs and the OS It allowed me to identify and highlight the consistencies and differences between the OS and the selected AWE. Initially, the data revealed that the coding for each set of feedback varied according to the software that generated this feedback. In addition, I found that the distribution of these codes was not consistent throughout all the documents, in some instances certain codes occur more frequently than others, while others did not occur at all; there were even more differences between those programs and texts that did not have any occurrence of a specific code.

WriteLab and PEG Coding

Table 2 represents my initial findings from the coding of the feedback generated by WriteLab and PEG. It was not possible to code any feedback from Eli Review and WriterKey because these programs did not generate automated feedback. The information in Table 2 reveals that the occurrence of codes varied according to which software generated the feedback. In addition, several of the OS codes did not occur in any of the two sets of feedback; there was relative consistency regarding which codes were not present in both documents. The frequency with which each code occurred within each set of feedback however, varied between WriteLab and PEG.

- The bulk of the feedback generated by PEG was coded to P7 and KC 1
- The feedback generated by WriteLab was primarily coded to RK 3 and P1.

41

Table 2 Results of WriteLab and PEG Coding

Coding	WriteLab	PEG
RK 1	79	29
RK 2	0	0
RK 3	340	66
RK 4	0	0
RK 5	0	0
CT 1	0	5
CT 2	1	0
CT 3	6	0
CT 4	1	0
CT 5	0	0
CT 6	0	0
CT 7	0	0
P 1	262	55
P 2	191	67
P 3	3	2
P 4	0	0
P 5	0	4
P 6	0	0
P 7	0	75
KC 1	65	99
KC 2	1	40
KC 3	0	0
KC 4	0	9
KC 5	0	0
KC 6	0	0

Instructional Documents Coding

I analyzed each instructional document individually, since every AWE adopted a different approach towards providing feedback. My examination revealed that there was slightly less consistency in the occurrence of codes across these 4 documents compared to the analysis of the feedback. In addition, the frequency of codes in the WriteLab and PEG

instructional documents were also not consistent with the frequency I identified after coding the automated feedback.

- P2 and P5 were the most commonly occurring codes in the instructional document taken from the PEG website.

- P2, RK3 and P1 had the highest level of frequency in the documents retrieved from the WriteLab website.

- P2 and P5 were the most frequently occurring codes in both the Eli Review and Writerkey documents.

Table 3 Results of Instructional Document Coding

Coding	PEG Doc	W.Lab Doc	Eli Doc	W.Key Doc
RK 1	0	3	0	2
RK 2	2	0	0	1
RK 3	0	11	2	2
RK 4	0	0	0	0
RK 5	0	0	0	0
CT 1	0	1	10	3
CT 2	1	0	0	2
CT 3	0	0	0	0
CT 4	0	3	0	0
CT 5	0	0	1	2
CT 6	0	0	0	2
CT 7	0	0	0	0
P 1	4	11	18	7
P 2	6	14	4	3
P 3	3	0	8	2
P 4	2	4	10	4
P 5	6	8	20	6
P 6	0	0	0	0
P 7	1	5	8	2
KC 1	0	2	0	0
KC 2	1	1	0	0

KC 3	0	0	0	0
KC 4	0	0	1	0
KC 5	0	0	0	0
KC 6	0	0	0	0

Similarities in the Coding

Despite the variations in the way that these software function, as well as the obviously different rhetorical purposes between the feedback and the instructional documents, I identified several consistencies across the different sets of data. The most significant of which was that 7 of the codes did not occur at all through the coding process; KC5, KC6, KC3, P6, CT 7, RK 4 and RK 5 were not present in the final set of coding. This consistency was likely because most of these codes were related to outcomes that addressed activities that needed to take place outside of the actual process of composition; the omitted codes deal with concepts like discipline or medium and how these concepts impacted the process of composition. It would be logistically difficult for programs concentrating on text related feedback to consider these issues since they depend on conditions that exist outside of the written document.

Another consistency between the feedback and instructional documents was the frequent occurrence of the P codes.

- P1 was one of the most commonly occurring codes within all the instructional documents, occurring a total of 40 times. It also occurs 262 times within the WriteLab feedback and 55 times in the PEG Feedback.
- P1 and P2 occur frequently and consistently throughout all 6 sets of coded texts

- P5 also occurred 40 times within the coding of all the instructional documents. This code occurs 5 times in the PEG Feedback. It did not occur within the WriteLab feedback.

- The P1 and P2 codes were present across all the data sets in a statistically significant manner. This consistency is in stark contrast to the way that the rest of the codes were identified across the sets of data. Finally, all the coded documents reflected at least some of the codes from each of the categories, further reflecting some degree of similarity in the relationships between the AWEs and the OS.

- Further comparison of the coding related to the instructional documents and the 2 sets of feedback also revealed some similarities in the relationship between these different sets of data. Coding for the software documents consisted mostly of the P codes; the two most common codes were P1 (19.4%) and P5 (18.9%). These codes are also evident in the other two sets of information, although they are present in different proportions and WriteLab Feedback coding does not contain P5. P4 was the only exception to this rule and only occurred in the coding of the instructional documents.

The similarity between the software documents and the coded feedback suggests a level of consistency between the information presented in the instructional texts and the actual feedback generated by the AWEs. I only identified 3 codes in the instructional documents that were not also identified in the coding of the feedback from WriteLab and PEG: P4, CT5 and RK 2. These three codes relate to genre, discipline and collaboration; they reflect examples of outcomes that are represented in the way that some of the

software function but that are not related to feedback about specific text. This relative consistency suggests that the instructional documents accurately represent the software they are discussing, which reinforces the appropriateness of the overall comparison being performed within the research. In addition, it also reflects a similarity between the kinds of tasks AWE are currently capable of performing as well as highlights outcomes to which these software are not able to contribute.

Differences in the Coding

I also identified several inconsistencies across the different sets of coded information. The number of times that each code occurred varied significantly depending on which software generated the feedback and further changed when I compared the codes and their frequency with the instructional documents.

- P7 is an example of this discrepancy; it occurs 75 times in the PEG Feedback but does not occur at all in the coding of the WriteLab Feedback. The frequency also varies within the instructional documents; occurring as often as 8 times in the Eli Review instructional documents while only being present once in the PEG Feedback.

- CT2 also occurred once in the WriteLab feedback but was not identified at all in the PEG Feedback. I identified more significant variations across the instructional documents where it was coded once in the PEG documents and twice the documents related to WriterKey while not being present in the Eli Review or WriteLab documents.

These variations in emphasis are related to the differences in the approach that each software uses to provide feedback as well as the different rhetorical goals of the instructional documents and the software generated feedback. Ultimately, while I can use these differences in the rate of occurrence for each code to provide information about how the different AWE interact with specific outcomes in the OS, the overall consistency in the presence of the codes across the different pieces of software and instructional documents is much more important as being reflective of the larger, generally positive relationship between these AWE and OS.

DISCUSSION

To complete this analysis, I operationalized the OS as a set of concise academic principles and an objective indicator of some of the perspectives of the academic community regarding FYC. I used the outcomes in the OS as a means of deconstructing the feedback generated by WriteLab and PEG as well as the instructional documents related to all 4 programs. I operationalized the outcomes in the OS as a means of accomplishing this analysis; this process was further made possible by the OS' necessarily non-specific and goal oriented qualities which allowed for it be analyzed and for broader meanings to be "unpacked" (Rhodes et al 2005, p.16).

Since each of the codes represent an outcome within the OS, I was able to draw conclusions about how the selected AWE reflect these outcomes as well as make some limited statements regarding the general nature of AWE software. My discussion also deals with the consequences of the typical pedagogical approach employed by these software, which emphasizes revision and feedback as well as how this approach is related

47

to the OS. I also identify some limitations to these software, some of which are logistical

inabilities that are most likely common to all AWE. I conclude by discussing the

relevance of this study to discussions about AWE and FYC as well as by suggesting ways

in which this research could be expanded to further analyze the relationships between

AWE and their role within the instruction of FYC.

Similarities between AWE and OS

My analysis concentrated on two different kinds of AWEs: those that generated

automatic feedback and those that used their automated capabilities to accommodate and

streamline the process of giving feedback. A comparison of these programs revealed that

they shared an approach in how they accommodated the feedback process.

Revision and Drafting. While each software facilitated feedback in different

ways by concentrating on different aspects of the OS, the consistent approach employed

by all the AWEs was an emphasis on revision and the importance of completing multiple

drafts. This approach was further reinforced by the consistent recurrence of the P1 and P2

codes within the automated feedback as well as within the instructional documents; these

outcomes are specifically related to drafting and other ideas associated with the process

of revision.

The AWE encourage revision in many ways that are consistent with the definition

of the P1 and P2 codes, including the suggestion to consider adding variety to writing, as

shown in the below example from the PEG feedback. This feedback was coded to both

P1 and P2 because it encouraged a revision that has many possibilities. Although the

statement is limited to one sentence, the frequency of these kinds of statements is

suggestive of the general emphasis placed on revision and drafting.

"Try adding strong verbs, specific nouns, adjectives, and adverbs."

WriteLab also accomplishes this same emphasis on rewriting by suggesting students rethink their word choices. This example was also coded as both P1 and P2 because it encourages revision but it is also phrased in such a way that it could lead the student to complete multiple drafts by considering the many different options for how this sentence could be rewritten:

"is when" Is is when necessary here? If not, remove it or replace it with in."

The instructional documents are able to more directly emphasize their focus on the process of revision and the way that these software operate under the assumption that recursive activities are important to how students learn to write. Sentences coded to P1 and P7 were those that more directly addressed these issues. In the Eli Review instructional texts, the below sentence was coded as both P1 and P2,

"You can coordinate write-review-revise cycles to encourage more timely feedback and revision."

Writerkey demonstrates the same pedagogical preference in the way that its instructional text explains how it works, this example was also coded to P1 and P2.

"A side-by-side view of their writing with comments helps students engage fully in the revision process."

This emphasis on Process based writing pedagogy is a unifying theme across the coding of the feedback and instructional texts. Consequently, this approach, which emphasizes revision as a means of improving the document, is also commonly accepted within different academic environments (Barnhisel, Stoddard and Gorman, 2012). In addition, while Process based writing can easily be attributed to many of the outcomes within the 'Process' category of the OS, it is also reflected in other outcomes from the

49

Rhetorical Knowledge and the Knowledge of Conventions categories; some of these outcomes are related to changing the structure or content of composition.

Common Limitations. My analysis also identified several practical limitations related to the AWEs ability to contribute to all the outcomes listed in the OS. These limitations are likely most closely related to the fact that all the AWE selected for this study provided feedback in a manner that was limited to standalone texts. There was therefore no opportunity to give feedback on any outcome related to more abstract concepts like context or genre, because that was beyond the logistical capabilities of these programs. Despite the flexibility with which the outcomes of the OS were intended to be used, several of them were still related to intangible concepts like discipline and reflection; these outcomes subsequently require activities that are beyond interacting with the text being developed, such as reading and developing an understanding of related concepts. In addition, any feedback related to secondary considerations like considering the medium within which the writing is taking place as well as determining the appropriateness of this medium were also beyond the scope of this study and likely beyond the capacity of programs that concentrate on individual texts and clauses. Similarly, other cognitive tasks such as considering genre is also something that cannot be accomplished by these programs. These AWE would only be able to assist with a specific aspect of this process or would require some outside intervention, possibly in the form of an instructor, as a means of achieving these larger, more summative outcomes.

This limitation remained consistent even after my examination of the instructional materials related to those programs that assisted with human feedback; interpreting this information however, requires a more nuanced perspective. The analysis of Writer Key

50

and Eli Review was obviously not as intensive as the process of analyzing the automated

feedback. While the coding maintained that helping students develop knowledge related

to context and genre is not something that the programs even claim to be able to

accomplish on their own, these results are complicated by the fact that these programs are

designed to be used along with outside instruction. Therefore, while, none of the

instructional documents programs discussed the possibility of providing feedback related

to summarizing the idea of a text or other more abstract activities, it is not possible to

assess whether this omission is related to a functional inability of the software or a failure

of the documents to properly explain their capabilities. The only takeaway from this

coding therefore, must be the fact that ideas related to these outcomes were not given

emphasis within these documents which is suggestive of the role this information plays

within the way that AWE operate.

Another universal limitation of all the selected AWEs was the fact that none of

them could address issues related to citation. There are obvious formatting and design

limitations associated with the way that my study was developed, in addition the personal

narrative style of the sample documents, meant that they would not necessarily include

any citations. The sample document labeled AW did however, include a 'works cited

page.' The use of these keywords as well as the traditional placement of this section at

the bottom of the page did not trigger any feedback related to citation. While the

programs provided some feedback encouraging the use of evidence and making logical

arguments, the software were not specifically designed to identify and address issues

related to composition. This limitation was supported by the fact that none of the

instructional documents included any references to citation, even while similarly

51

addressing related concepts from the Critical Thinking category including logic and how to develop strong arguments.

Shared Goals. My review of the instructional documents associated with each AWE also allowed me to compare the way that the 4 texts discuss their features with the outcomes identified as best practices by the academics who developed the OS. The conclusions from this review are consistent with the results of the coding and highlighted the fact that each software emphasized the role of revision and feedback in the way that it operated. It also reinforced the finding that certain outcomes were universally absent from how these programs functioned and therefore not consistent with their stated purpose.

WriterKey outlines its main features as "Draft differently. Comment easily. See Revision at work." This section clearly emphasizes revision and review as being fundamental to the way that WriterKey works; it highlights its ability to facilitate these two activities for both teachers and students. Similarly, in "How Eli Review Works", the emphasis is also placed on 'write-review-revise cycles,' which is a reference to the program's focus on developing and building on small writing activities. Similarly, WriteLabs' "How it Works" section is immediately separated under the headings of "Write. Revise. Review. Repeat." While "About PEG Writing Scholar" identifies these same activities in its explanation of how PEG works, the only difference being that this document also addresses scoring which is not a stated goal of any of the other programs, since they do not include this feature.

The close relationship between the expressed purposes of each of these programs further supports the results of the coding, which highlights commonalities between the 2

different kinds of feedback and the 4 instructional documents. Just like the coding, my review of the goals of each program reveals a consistent pedagogical approach towards teaching as well as an emphasis on specific kinds of feedback (focused on recursivity). In addition, my analysis of this information similarly highlighted a failure to address any goals related to reading, technology, the writing medium or any other activities that were not directly related to the processed text.

Relationship between AWE and the OS

My examination of the specific occurrence of the codes within the instructional texts and the automated feedback is useful in determining exactly which outcome the program is emphasizing in each instance; however, the overall relevance of this quantitative data may not be that significant when discussing the larger relationship between AWEs and the OS. The OS' focus on outcomes as opposed to standards means that this document "provides curricular parameters without articulating specific levels of proficiency" (Olsen, 2012, p.19). With the exception of Rhetorical Knowledge, the OS does not identify any one outcome as being the most important or try to control what kind of attention should be paid to any specific area of knowledge. Therefore, the fact that my research revealed a significant amount of this feedback is geared towards encouraging students to reach many of the different outcomes in the OS, may be more significant than the number of times that a single outcomes is addressed by a specific AWE .

This idea of outcomes versus standards is important to the consideration of how the categories of the OS are represented within the coding. Within each larger category, there is a very close relationship between each specific outcome; these outcomes may have meanings that overlap in a manner that complicates the coding and therefore it is

most effective to consider them as representative of a collection of ideas instead of as standalone concepts. An example of this close similarity is that the first two outcomes in the Rhetorical Knowledge category, RK1 and RK2 both deal with the idea of revision to suit a rhetorical purpose; the qualifying difference between the two is that RK2 emphasizes how to use genre while RK 1 focuses on changes to the composition. Given this perspective, it is more useful to examine the relationship between the coded texts and the larger OS categories as a means of understanding the relationship with the AWE.

Rhetorical Knowledge. My Literature review highlights the fact that the role of Rhetorical Knowledge was given special emphasis within the OS as well as by academics who discussed FYC. Maid and D'Angelo (2012) refer to Rhetorical Knowledge as the "uber outcome" (p.257) because it is the outcome upon which all other outcomes are dependent. This kind of emphasis suggests that if the AWEs are to be effective at representing the outcomes in the OS, then they must encourage and develop the different stages of Rhetorical Knowledge. Ultimately, most of the outcomes associated with Rhetorical Knowledge were all identified and coded throughout the texts.

According to my coding, the 4 AWEs all provided or claimed to provide feedback that encouraged thinking about the rhetorical situation within which the writing was taking place. The Rhetorical codes identified in the different texts were: RK1 which was related to editing writing to adjust to the rhetorical situation. RK 2 related to using genre to adjust to the rhetorical situation and RK 3 which is related to how the writing could be understood or interpreted. The only codes that were not identified were RK 4 and RK 5 which both deal with commenting on and discussing the appropriateness of the composition medium and was therefore not able to accomplished by these programs that

54

are intended to analyze text. The fact that Rhetorical Knowledge is uniformly represented throughout the coding suggests that this outcome is important to the way that these programs accomplish the goal of providing feedback.

The outcomes in the Rhetorical Knowledge category are not the most popular throughout the coding but they are represented uniformly in both the feedback and instructional documents. The different sentences and pieces of feedback coded to this category generally encouraged students to reconsider the decisions they have made while composing by thinking about how it would be understood by readers. The different programs accomplish this general task in different ways; in the instructional documents, I applied the coding to statements that discussing how students are encouraged to write for different purposes, which in this example from the PEG document is related to different assignments.

"Instructor-recommended prompts make differentiating assignments an easy task."
I coded the automated feedback to the Rhetorical Knowledge category when the feedback encouraged students to consider how their writing could be read by others. In this example feedback generated by WriteLab, this is accomplished by encouraging consideration of how to simplify the writing,

""some of the" Here is an opportunity to simplify your writing. Consider saying some instead of some of the."
In addition, this kind of feedback seems to be generated in such a manner that encourages students to consider revision and rewriting for the purpose of meeting the different rhetorical goals of their writing. This approach reinforces the relevance of this specific outcome by connecting it to another academic conversation, the importance of

revision, already identified as being important by academics like Daneshvar and Rahimi (2014) and Semke (1984).

Critical thinking, Reading and Composing. Coding related to outcomes in the Critical Thinking category occurred very rarely within the automated generated feedback. The main reason for this exclusion is the fact that several of these outcomes are not something that a software program analyzing individual pieces of text would be able to accomplish; these outcomes reference different activities like reading and researching which means that it is immediately excluded any kind of coding related to text-based analysis. The codes that were identified within the text were those that encouraged the idea of reading within the process of delivering the feedback. In this example of feedback generated by PEG, the text encourages rereading the sentence to develop an understanding of its meaning and was coded to CT2.

"This sentence appears to be a run-on. You may need a period or some other end punctuation following "with the friendships". If you think the sentence is not a run-on, then there should be some other punctuation here, like a comma or a dash, or, if what follows is a direct quotation, a comma and quotation marks. If none of these suggestions seem right, make sure you've spelled all the words in your sentence correctly, i.e. you haven't used a word incorrectly, or that you haven't left out a word or added an unnecessary word."

While feedback associated with this category was not consistently present throughout the two sets feedback, the instructional documents contained more consistent examples of these outcomes. The discrepancy between these two texts was likely related to the fact that as instructional texts, these documents were not necessarily limited by logistics in the way that they were able to discuss the process of feedback and the things that they hoped to accomplish. My analysis of these instructional documents did not identify any significant discrepancy between the texts related to the automated feedback

generating programs and those related to the programs that accommodated human feedback; this consistency reinforces the idea that it is the format of the documents and not the pedagogical approach of the software that impacted the variations in coding.

In the example from Eli Review, this excerpt discusses how the program allows instructors to create reviews which can be read by students as a means of helping them develop feedback and subsequently develop their knowledge This feedback was also coded to CT2.

"You can create reviews that function like surveys to guide reviewers, focusing their feedback on the important criteria for learning."

In general, the infrequent occurrence of these codes throughout the automated feedback as well as the relatively sparse presence of these codes within the instructional documents suggests that there is an inherent incompatibility between this category of the OS and the way that these AWE function and develop feedback.

Processes. Outcomes from the Processes category were the most frequently occurring in both the feedback and the instructional documents. This was likely a result of the fact that unlike the Critical Thinking and Reading outcomes, outcomes associated with Process were very clearly related to what AWEs were trying to accomplish. These outcomes emphasize feedback, collaboration and revision. Like the outcomes associated with Rhetorical Knowledge, the outcomes in the Process category are grounded in academic discourse that highlights revision and rewriting as an effective means of improving student writing. In this example from WriteLab, coded to P1, it is clear how the feedback provided examples of word choice intended encourage students to consider the different ways that their text could be revised,

"*"Basically" Delete Basically or replace it with Mostly, Mainly, or Chiefly".*

The instructional documents contained even more evidence of Process based outcomes. In addition to the emphasis on revision, the themes of feedback and collaboration are also obviously central to how that these programs operate. In one example from the WriterKey overview, there are references to both the importance of feedback as well as references to collaboration by mentioning the role of comments. This sentence was coded to P1 and P2.

"*A side-by-side view of their writing with comments helps students engage fully in the revision process.*"

Similar however to the automated feedback, the instructional documents also did not contain any references to the outcome related to learning different kinds of technology. The singular absence of this outcome implies that this is either an approach that is not considered relevant by developers of these programs or something that they are not capable of doing; students who use these AWE are obviously immediately restricted to a single type of technology since these programs are limited to analyzing a specific kind of text and cannot identify different kinds of formatting.

Knowledge of Conventions. This outcome category is the least represented within the coding of the automated feedback. Outcomes related to Knowledge of Conventions deal with issues of grammar, paragraph structure, genre, design and citation. Grammar is obviously something that software can identify and analyze and it is also one of the only outcomes from this category that is consistently identified throughout the coding of the feedback. This was generally in the form of generic corrections related to

58

spelling, punctuation or an explanation of grammar rules like below in the example from WriteLab which was coded to KC1,

"*Instead of using the passive voice with disbelief my teacher made when I finished, try converting your verb into the active voice by specifying who made my teacher when I finished.*"

The other outcome in this category occurred most frequently within the feedback generated by PEG and was specifically related to paragraph structure and organization. These kinds of feedback, as seen in the next example coded to KC2 (as well as RK1 and RK3), commented on the writing by discussing how the different elements were working together.

"*You have a beginning and an ending, but try to make your beginning grab your reader's attention and your ending more interesting.*"

Except for these outcomes and some minor references to character spacing, which was coded as design or structure, there were no other references to any outcomes in this category.

The instructional documents contained even fewer references to this outcome. One explanation for this could be that the programs attempted to make claims that were more sophisticated than simple grammar corrections and therefore concentrated on discussing their more advanced capabilities. The references that were included were rare and usually discussed in relation to other features of the program, like in the example below from the "How Eli Review works" section coded to KC4.

"*Whether students are composing writing-to-learn responses or formal drafts, Eli Review helps students get more from peer learning.*"

None of the instructional documents discussed genre, design or convention; this reinforces the fact that these principles are not significantly addressed by the automated

feedback and suggests that these functions are not something that these AWE can perform.

AWEs and Feedback

The Writelab and PEG feedback were mostly coded to Rhetorical Knowledge and other outcomes associated revision, while the instructional documents that discussed the how these programs operate were similar coded to more general outcomes associated with feedback and collaboration. This approach implies, not surprisingly, that these programs reflect the academic emphasis on feedback and collaboration discussed by scholars like Elham Daneshvar and Ali Rahimi (2014). However, these scholars, along with others in the field are also concerned with how feedback is given and with finding a way of ensuring that it is done effectively. This approach is addressing concerns echoed by scholars like Dew (2012) who worries that "students as developing writer do not asses and respond to peer writing as well as professors" (p.13) and even before that by Semke who states that most feedback is ineffective and teachers are actually wasting their time by concentrating on this activity within the classroom. Therefore, while the larger concentration on feedback evident throughout the coding of the AWE represent the way in which these programs are designed to work, the supplemental emphasis on outcomes associated with rhetorical knowledge, revision and collaboration represent a pedagogical preference about how these programs seek to deliver and encourage this feedback.

By examining the consistencies in the way that each of these software provide feedback, I am able to identify several pedagogical similarities. All AWEs also favor an approach of giving feedback that emphasizes peer review and how this helps to develop student writing. This approach is once reflective of the concepts associated with Process

based writing pedagogy and appears to be consistent with how all the AWEs are designed. Each of these programs provide students with feedback throughout the writing, ultimately guiding them towards proficiency which is generally measured by some kind of a score. WriteLab does not have a grading component, however, if it is used within classroom environments then it can be implied that the writing knowledge will eventually be assessed in some way. PEG Writing Scholar provides a grade but allows students to make adjustments to their writing as many times as needed before submitting the paper for a final score. Eli Review and Writerkey are also built around using feedback from peers and teachers as a means of helping students improve their writing before it is completed and submitted.

CONCLUSION

The results of my research suggested that these AWE primarily develop feedback that is concerned with encouraging revision and drafting throughout the writing process. In addition to providing insight into a common approach employed by these programs, this information also suggests a pedagogical preference. This preference could be valuable to the way that instructors choose to use these programs within classroom environments. Several of the outcomes in the OS as well as supplemental research by academics like Barnhisel, Stoddard and Gorman (2012) and Hillocks Jr. (2007) suggest that an emphasis on revision is a legitimate approach towards teaching writing. Instructors could consider this focus on revision when determining how to include these programs into their lessons; AWE could provide them with an expedient approach towards accomplishing the well-recognized important task of encouraging students to

think about their writing as an iterative process; this would allow them to concentrate on providing feedback in other areas that are also important to their instructional goals.

My research further addressed the idea of expediency by highlighting those outcomes with which the AWE do not have a strong relationship. By recognizing those things that these programs can accomplish and those that they cannot, instructors could use these programs as guides and remain confident in the fact that the non-text related feedback related to things like medium, technology and discipline are still being addressed. This idea of expediency is important because it is the chief reason used to criticize feedback by academics like Semke (1984) and is the reason why other scholars like Daneshvar and Rahimi (2014) and Hillocks Jr. (2007) are concerned with improving the feedback process, they recognize that it can dramatically impact the experiences of teachers and students.

Future studies could expand on my research by adopting a more inclusive methodological approach that also examines human feedback accommodated and guided by AWE designed for this purpose. This analysis would allow for greater comparison of the way that the two-different families of AWE approach the feedback process. In addition, by incorporating actual examples of these kinds of feedback, it would be possible to further determine whether the fact that these programs include a human component affects how it can accomplish the outcomes in the OS. Such a study would directly address any inconsistences that may exists between the instructional documents and the actual feedback.

Finally, additional research could be done to add an evaluative component to this research. This would include an examination of whether the feedback was accurate when

compared to the content of the sample documents. By examining the accuracy of feedback, this kind of research would be able to make more concrete assessments regarding how AWE were actually executing their capabilities, which is important in order to determine if these programs deserve as central a role within classroom instruction as initially suggested by my research. This research may result in additional outcomes being excluded from the coding of the AWE and subsequently refine the interpretation of how these programs can be used.

While there is obviously an opportunity to expand and refine my study in the future, I also make two contributions to the existing AWE research. I provide a foundation for understanding the general capabilities of AWE; this information can then be used to contribute to the existing discussion about these programs and their role within classroom environments. Most significantly however, by using the outcomes as a basis for coding, I introduce a heuristic with which to examine AWE for possible use in FYC.

REFERENCES

About PEG Writing Scholar. (n.d.). Retrieved from
http://www.pegwritingscholar.com/about

Barnhisel, G., Stoddard E., & Gorman J.. (2012). Incorporating Process-Based Writing
Pedagogy into First-Year Learning Communities: Strategies and Outcomes. *The Journal
of General Education, 61*(4), 461-487. doi:10.5325/jgeneeduc.61.4.0461

Bowie, J.L., & McGovern, H.A. (2013). Decoding our scholarship: The state of research
in computers and writing from 2003-2008. *Computers and Composition: An
International Journal for Teachers of Writing,* 30(3), 242.
Doi:10.1016/j.compcom.2013.07.001

Brearey, Oliver: "Understanding the Relationship between First- and Second-Semester
College Writing Courses" Teaching English in the Two Year College (42:3) Mar 2015,
244-263. (2015)

Burstein, J., Elliot, N., & Molloy, H. (2015). Informing Automated Writing Evaluation
Using the Lens of Genre: Two Studies. *CALICO Journal*, 33(1). Retrieved 16 January,
2017, from https://journals.equinoxpub.com/index.php/CALICO/article/view/26374

Chapelle, C. A., Cotos, E., & Lee, J. (2015). Validity arguments for diagnostic
assessment using automated writing evaluation. Language Testing, 32(3), 385-405.
doi:10.1177/0265532214565386

Chandrasegaran, A., Ellis, M., & Poedjosoedarmo, G. (2005). Essay assist-developing
software for writing skills improvement in partnership with students. *Regional Language
Centre Journal,* 36(2), 137-155. Doi:10.1177/0033688205055567

Chen, Chi-Fen Emily (2008). "Beyond the Design of Automated Writing Evaluation:
Pedagogical Practices and Perceived Learning Effectiveness in EFL Writing Classes".
Language learning & technology (1094-3501), 12 (2), p. 94.

Cope, B., Kalantzis, M., McCarthey, S., Vojak, C., & Kline, S. (2011). Technology-
mediated writing assessments: Principles and processes. *Computers and Composition,
28*(2), 79-96. doi:10.1016/j.compcom.2011.04.007

Daneshvar, E., & Rahimi, A. (2014). Written corrective feedback and teaching grammar.
Procedia - Social and Behavioral Sciences, 136, 217-221.
doi:10.1016/j.sbspro.2014.05.317

Dando, P. (2016). Traditional Literacy and Critical Thinking. Knowledge Quest, 44(5), 8-
12.

Dew, D. (2012). CWPA Outcomes Statement as Heuristic for inventing Writing-About-
Writing Curricula. In Behm, N. et al (Eds.), *The WPA Outcomes Statement*: *A Decade
later.* (p.3-17). Anderson, South Carolina: Parlor Press.

El Ebyary, K. (2010). The impact of computer-based feedback on students' written work. International Journal of English Studies, 10(2), 121-145. Retrieved from http://login.ezproxy1.lib.asu.edu/login?url=http://search.proquest.com.ezproxy1.lib.asu.edu/docview/871381485?accountid=4485

Farrell, T. B. (1976). Knowledge, Consensus, And Rhetorical Theory. Quarterly Journal Of Speech, 62(1).

Gallagher, C. (2012). The Trouble with Outcomes: Pragmatic Inquiry and Educational Aims. College English, 75(1), 42-60. Retrieved from http://www.jstor.org.ezproxy1.lib.asu.edu/stable/24238306

Gallagher,E. (1999). Errors and expectations: A guide for the teacher of basic writing. *Teaching English in the Two Year College*, 27(1), 118.

Gorlewski, J., Krickovich, S. R., & Gorlewski, D. A. (2011). Research for the classroom: Review of "narrative writing: Learning a new model for teaching" by george hillocks jr. The English Journal, 100(3), 109-111.

Griffin, J.& Minter, D. : "The Rise of the Online Writing Classroom: Reflecting on the Material Conditions of College Composition Teaching" College Composition and Communication (65:1) Sep 2013, 140-161. (SPECIAL ISSUE: The Profession) (2013)

Harrington, S., Malencyzk, R., Peckham, I., Rhodes, K., & Yancey, K. B. (2001). WPA outcomes statement for first-year composition. *College English, 63*(3), 321.

Harrington, Susanmarie; Rhodes, Keith; Fischer, Ruth; and Malenczyk, Rita. (2005) "The Outcomes Book: Debate and Consensus after the WPA Outcomes Statement". All USU Press Publications. Book 154. http://digitalcommons.usu.edu/usupress_pubs/154

Hewett, Beth L. "A review of WriteLab." WLN: A Journal of Writing Center Scholarship, vol. 40, no. 3-4, 2015, p. 8+. Academic OneFile, Accessed 6 Feb. 2017.

Horner, B., & Lu, M. (2010). Working Rhetoric and Composition. College English, 72(5), 470-494. Retrieved from http://www.jstor.org.ezproxy1.lib.asu.edu/stable/20749293

How Eli Review Works. (n.d.). Retrieved from http://elireview.com/learn/how/

How it Works. (n.d.). Retrieved from https://www.writelab.com/how-it-works/

Knievel, M. (2009). What is humanistic about computers and writing? Historical patterns and contemporary possibilities for the field. *Computers and Composition, 26(2), 92-106.* Doi: 10.1016/j.compcom.2009.02.002

Lowe, K. (2006). S. Harrington, K. Rhodes, R. Fischer and R. Malenczyk.(eds). *The Outcomes book: Debate and consensus after the WPA outcomes statement*. Utah State

University Press: Logan, UT (2005) ISBN 0874216044 240 pp. *Assessing Writing, 11*(3), 224-229. doi:10.1016/j.asw.2006.05.003

Maid, B. & D'Angelo, B. (2005). Is Rhetorical Knowledge the Uber Outcome. In Behm, N. et al (Eds.), *The WPA Outcomes Statement*: *A Decade later.* (p.257-270). Anderson, South Carolina: Parlor Press.

Miller, G. (2009). Computers as Writing Instructors. *Science,* 32(5910), 59-60. http://www.jstor.org.ezproxy1.lib.asu.edu/stable/20177117

Miller, M. D. (2014). *Minds online: Teaching effectively with technology*. Cambridge, Massachusetts: Harvard University Press.

Moore, N. S., & MacArthur, C. A. (2016). Student use of automated essay evaluation technology during revision. *Journal of Writing Research, 8*(1), 149-175. doi:10.17239/jowr-2016.08.01.05

Mulnix, J. W. (2012). Thinking Critically about Critical Thinking. Educational Philosophy & Theory, 44(5), 464-479. doi:10.1111/j.1469-5812.2010.00673.

Olson, W. (2012). The Politics of Pedagogy: The Outcomes Statement and Basic Writing. In Behm, N. et al (Eds.), *The WPA Outcomes Statement: A Decade later.* (p.18-36). Anderson, South Carolina: Parlor Press.

Patterson,J.P., and David Duer. (2006). High School Teaching and College Expectations in Writing and Reading. *The English Journal,* 95(3), 81-87. http://doi.org.ezproxy1.lib.asu.edu/10.2307/30047049

Ranalli, J., Link, S., & Chukharev-Hudilainen, E. (2017). Automated writing evaluation for formative assessment of second language writing: Investigating the accuracy and usefulness of feedback as part of argument-based validation. *Educational Psychology, 37*(1), 8-25. doi:10.1080/01443410.2015.1136407

Rankins-Robertson, S. (2012). The Outcomes Statement as Support for Teacher Creativity: Applying the WPA OS to Develop Assignments. In Behm, N. et al (Eds.), *The WPA Outcomes Statement: A Decade later.* (p.58-70). Anderson, South Carolina: Parlor Press.

Rudes,D.,Portillo,S., Sloas,L. & Hutzell,K.(2014) Students as Scholars & Writers: Teaching & Assessing Undergraduate Writing in Capstone Course. *Journal of Criminal Justice Education* 25(1), 54-68. Doi: http://dx.doi.org/10.1080/10511253.2013.849353

Scruggs, T. (1988). Psychology of Computer Use: VI Effectiveness of Computer Assisted Instruction in Expository Writing. *Perceptual and Motor Skills: Volume 67*, Issue 3, pp.871-877. Doi: 10.2466/pms.1988.67.3.871

Semke, H. D. (1984). Effects of the red pen. Foreign Language Annals, 17(3), 195. Retrieved from http://login.ezproxy1.lib.asu.edu/login?url=http://search.proquest.com.ezproxy1.lib.asu.edu/docview/1311687782?accountid=4485

Shun, F.(2014). The application of schema theory in teaching college English writing. *Theory and Practice in Language Studies,* 4(7), 1476. Doi:10.4304/tpls.4.7.1476-1482

Steve Graham, Karin Sandmel. (2011) The Process Writing Approach: A Meta-analysis. The Journal of Educational Research 104:6, pages 396-407.

Stevenson, Marie (2016). "A Critical Interpretative Synthesis: The Integration of Automated Writing Evaluation into Classroom Writing Instruction." *Computers and Composition,* 42, 1-16. Doi:http://dx.doi.org/10.1016/j.compcom.2016.05.001

Stevenson, M., & Phakiti, A. (2013;2014;). The effects of computer-generated feedback on the quality of writing. *Assessing Writing, 19*, 51. doi:10.1016/j.asw.2013.11.007

Soles, D. (2002). A Comment on the "WPA Outcomes Statement for First-Year Composition" College English, 64(3), 377-378. doi:10.2307/3250740

Taavitsainen, I. (2001). Changing conventions of writing: The dynamics of genres, text types, and text traditions. *European Journal of English Studies, 5*(2), 139-150. doi:10.1076/ejes.5.2.139.7309

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Medical Research Methodology, 8(1), 45-45. doi:10.1186/1471-2288-8-45

Turnitin; the white house and department of education recognize turnitin revision assistant among other leaders seeking to improve how testing can inform instruction.(2016). Politics & Government Business, , 16. Retrieved from http://login.ezproxy1.lib.asu.edu/login?url=http://search.proquest.com.ezproxy1.lib.asu.edu/docview/1850884934?accountid=4485

Vitartas, P., Heath, J., Midford, S., Ong, K., Alahakoon, D. & Sullivan-Mort, G. (2016). Applications of Automatic Writing Evaluation to Guide the Understanding of Learning and Teaching. In S. Barker, S. Dawson, A. Pardo, & C. Colvin (Eds.), *Show Me The Learning. Proceedings ASCILITE 2016 Adelaide* (pp. 592-601).

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: Defining the classroom research agenda. *Language Teaching Research, 10*(2), 157-180. doi:10.1191/1362168806lr190oa

Ware, P.(2011). Computer Generated Feedback on Student Writing. *TESOL Quarterly,* 45(4), 769-774. Retrieved from http:www.jstor.org.ezproxy1.lib.asu.edu/stable/41307666

White, M. D. & Marsh, E. E. (2006). Content Analysis: A Flexible Methodology. *Library Trends* 55(1), 22-45. The Johns Hopkins University Press. Retrieved January 16, 2017, from Project MUSE database

Wiley, M. (2005). Outcomes are not Mandates for Standardization. In Harrington S. Rhodes K., Fischer R.& Rita M.(Eds.), *The Outcomes Book*: *Debate and Consensus after the WPA Outcomes Statement* (p.24-31). Logan, Utah: Utah State University Press.

Wilson,V. (2011). Evidence Based Library and Information Practice, 6(4), 177–179. Retrieved from
https://ejournals.library.ualberta.ca/index.php/EBLIP/article/view/12180/13124

WPA Outcomes Statement for First-Year Composition (2014, July 17). Retrieved from
http://wpacouncil.org/positions/outcomes.html

Writerkey Overview. (n.d.). Retrieved from
http://www.macmillanlearning.com/catalog/static/bfw/WriterKey-Tour/home.html

Yancey, S. (2005). Standards, Outcomes and All that Jazz. In Harrington S. Rhodes K., Fischer R.& Rita M.(Eds.), *The Outcomes Book*: *Debate and Consensus after the WPA Outcomes Statement* (p.18-23). Logan, Utah: Utah State University Press.