An Investigation of Outcomes for a Course Taught in Three Different Learning Environments

Mary Gozza-Cohen
Thomas Jefferson University, USA

Abstract

Nationally and regionally there are insufficient numbers of qualified special education teachers to meet current demand. Online course delivery has been proposed as one way to reach more students and increase the number of special education teachers. A meta-analysis conducted by the U.S. Department of Education reviewed the research on online learning and reported that on average, students in online learning conditions performed better than those in traditional classes [1]. However, among other issues noted, it was specified that many of the included studies did not control for curriculum materials and various aspects of pedagogy. This exploratory study investigated student outcomes in a graduate level special education course taught in three learning environments (traditional, online, and blended) by the same instructor using the same materials and activities for all three sections of the course.

1. Introduction

The shortage of special education teachers in K-12 settings has forced institutions of higher education to search for innovative ways to attract and retain students in special education teacher education programs. It is important to determine if the use of online learning environments can prepare special educators as effectively as traditional courses in ways that indicate both knowledge or skill acquisition, and deep or meaningful learning and understanding, beyond just course or test grades. In addition, if the blended approach is the most effective [1], does the blended model add enough to the online course to justify eliminating purely online options for students who cannot attend any campus classes? If online courses alone can prepare the students effectively, then this model may be one way to expand the population of potential special education teachers; offering flexibility in scheduling and geographic location may be a way for students who otherwise would not be able to attend campus classes to receive some of the preparation to teach special education students.

In the field of special education teacher preparation, the literature on online instruction is not extensive, but initial findings indicate positive outcomes for online instruction with limited outcome measures included in the studies [2, 3, 4, 5]. Since alternative delivery methods, such as online instruction, have been proposed as a way to address the chronic and increasing shortage of special education teachers available to meet the needs of students with disabilities in the nation's schools, this study adds to existing research by reporting an investigation of a graduate special education course taught in three different learning environments [6].

This study sought to address some of the previous methodological concerns about research on alternative learning environments in college and university courses by comparing student level of understanding of content, achievement, and opinions using multiple indicators, across three sections of a graduate-level special education course. All three sections were taught by the same seasoned professor using the same course materials, assignments, and content delivered in three different learning environments: traditional, blended, and online.

2. Research Design and Methods

This study involved a concurrent embedded strategy of mixed methods which is "identified by its use of one data collection phase, during which both quantitative and qualitative data are collected simultaneously…unlike the traditional triangulation model, a concurrent embedded approach has a primary method that guides the project and a secondary database that provides a supporting role in the procedure" [7]. The primary database in the current study is quantitative based on course grades and researcher-assigned scores on student work, and the secondary supporting database is qualitative based on students' open-ended responses to questions on an end-of-semester course survey. Using the qualitative data to explore the experiences of the students in the three different learning environments, this method was used to gain a broader perspective. It was anticipated that the addition of qualitative data would help to explain or support some of the results of the quantitative analyses.

1. Are there differences in level of understanding of course content among graduate students in a

special education course delivered in three different learning environments?
2. Are there differences in course grades among graduate students in a special education course delivered in three different learning environments?
3. Are there differences in opinions about the course structure among graduate students in a special education course delivered in three different learning environments?
4. Are there relationships among level of understanding, course grades, and student demographic data from graduate students in a special education course delivered in three different learning environments?

2.1. Participants

The participants for the study consisted of forty-one graduate students in their final semester of a two-year special education and literacy teacher education program in a large university in the Northeast. The students completed majors or minors in the liberal arts and sciences. The program is very competitive, accepting approximately one quarter to one third of applicants, so the students typically have strong grade point averages (GPA). The students were randomly assigned to the online and traditional sections, and the following year all students were assigned to the blended condition. Demographic data were collected from student records on file with the Special Education program and used to ascertain if there were differences in age, undergraduate GPA, and program application writing scores that might impact the results of this investigation. The means and standard deviations for age, undergraduate GPA and writing score are shown in Table 1. A one-way analysis of variance was used to analyze this demographic data and revealed no significant differences in the students in the three environments were found on the demographic data. The ANOVA data are shown in Table 2.

Table 1. Means and standard deviations for age, undergraduate GPA, and writing score

Learning Environment	Age		UGrad GPA		Writing Score	
	M	*SD*	*M*	*SD*	*M*	*SD*
Traditional (n=11)	29.18	(4.91)	3.36	(.23)	26.68	(2.94)
Online (n=11)	28.91	(3.80)	3.26	(.47)	26.86	(4.11)
Blended (n=20)	28.90	(6.21)	3.30	(.31)	26.80	(2.83)
Total (n=42)	28.98	(5.22)	3.31	(.39)	26.78	(3.15)

Table 2. Summary of analysis of variance by age, undergraduate GPA and writing score

Source	SS	*df*	MS	*F*
Age	.63	2	.31	0.01
Within/Error	1120.34	39	28.72	—
Total	1120.97	41	—	—
Undergraduate GPA	.05	2	.02	.218
Within/Error	4.67	39	.01	—
Total	4.72	41	—	—
Writing Score	.19	2	.09	.009
Within/Error	408.88	39	10.48	—
Total	409.07	41	—	—

($p < .05$)

2.2. Instructor

The senior professor in the special education teacher preparation program taught all three sections of this course as part of her regular course load. She is a tenured full professor who had taught this course for over 25 years in the traditional face-to-face format. She was recognized for her teaching by Excellence in Teaching Awards from both her university and the larger university system. Prior to teaching the online and blended formats, she received over a year of additional training and support to voluntarily convert the course into both a fully online and a blended course, and did so both willingly and enthusiastically.

2.3. Course description, materials and activities

The course presented theoretical positions, assessment techniques, planning procedures, and teaching methods relevant to preparing students with severe and multiple disabilities to meet the State Learning Standards. It focused on children and adults with handicapping conditions such as severe mental retardation, severe physical and neurological disabilities, severe behavioral and emotional disturbances, autism, multiple disabilities and severe sensory problems. It emphasized how to develop effective educational programs for this population based on the alternative learning standards, ecological assessments, functional assessments, Universal Design for Learning (UDL), and the use of all available resources, including technology, adaptive equipment, and the alternative assessments, in the least restrictive settings.

The course, regardless of format of presentation (face-to-face, online or blended), had the same instructor and used the same text, readings, videos, and assignments. In addition, all students had the full PowerPoint presentations that had been developed for the course. The course activities included integrated discussions, collaborative activities, and a

group project. Since the initial written responses to discussion prompts were examined in two different ways in this study, it is important to explain this discussion prompt response task in more detail.

Two integrated initial responses to two different discussion prompts, each a written integration of assigned course materials, were included in this study. Discussion Response One was on the history of special education and inclusion, and Discussion Response Two was on culture and genetics. Each section of the course had the same prompt for these initial discussion responses that required students to read materials from the text and assigned articles, watch digitized videos, explore websites, and if relevant, relate the materials to personal experiences. The students in the traditional section of the course watched some of the video material in class but had the same written assignment prompt as well as links to all the digitized videos and articles. The assignment required students to integrate most, if not all, of the materials into an initial written response to a discussion prompt; the online and blended sections of the class posted their initial responses online, and at the end of the class period, the traditional class turned in their typed initial responses (prepared before the class session) following the in-class small group discussion.

This study used archival data; copies of students' work on the integrated written responses to discussion prompts were retained online or collected in hard copy, and these artifacts were used, in addition to course grades and responses to a student survey, to explore if the instructional format impacted student performance.

3. Instruments

The Structure of Observed Learning Outcomes Taxonomy, instructor designed discussion rubric, course activities and end-of-semester survey are described in the sections that follow.

3.1. SOLO taxonomy

The Structure of the Observed Learning Outcomes (SOLO) Taxonomy that has been used for a number of years to systematically assess the levels of complexity of student understanding and learning and was used in this study to rate students' understanding of the course content as evidenced by their two individual written responses to the discussion prompts [8, 9, 10, 11, 12]. According to Biggs and Collis (1982), the SOLO taxonomy is not content-dependent and consists of five levels with each student artifact scored from 1 to 5 using the following hierarchy:

1. Prestructural – the student misses the point and shows understanding at the individual word level, repeats information, uses different words to say the same thing, does not add substance;
2. Unistructural – Meets only one part of the task and misses the other part(s), adds a little more on pre-structural, uses one source;
3. Multi-structural – Does not understand/ address the key issues, presents facts in an unstructured manner, knowledge telling with no connections, sees the trees but not the woods;
4. Relational – Student no longer lists facts, the paper addresses the point, does some conceptual work, but stays with what is given; and
5. Extended Abstract – Goes beyond what is given, uses higher level of abstraction, gives a perspective that changes what we think about the topic, takes the argument into a new dimension.

Biggs and Collis (1982) reported that the initial SOLO inter-rater reliability correlations ranged from .71 to .95. Construct validity for the SOLO was established in 1982 and was based on a correlation between the SOLO and other measures of achievement which suggested that "SOLO is closely involved with school achievement and high SOLO levels are obtained by highly intrinsically motivated students who search for meaning" (Biggs & Collis, 1982, p. 194).

3.2 Discussion rubric

The rubric was developed by the course instructor to evaluate initial written responses to the discussion prompts; the students in the traditional section were only evaluated on the quality component since the other two categories could not be monitored, or did not apply to the face to face class.

Criteria	Excellent 4 pts	Good 3 pts	Satisfactory 2 pts	Poor 1 pt
Quality of Starter Draft	Starter drafts are original, relevant, comprehensive, bring out major themes and issues, clearly written, demonstrate reading of all materials, open-ended to allow for ample discussion.	Starter drafts are relevant, related to major themes, demonstrate reading of most materials, are open-ended.	Starter drafts are related to themes, demonstrate reading of some materials, and allow for some on-going discussion.	Starter drafts are too specific to reflect major themes or are irrelevant, not clear what you read.
Partici-pation	Highly responsive making connections that help to expand on points,	Responsive to comments of others, making connections and	Somewhat responsive to others, making some connections and	Comments not related to discussion thread, postings missing.

	synthesize information and build accuracy.	expanding on points.	expanding on some points.	
Timeli-ness of Drafts	Has starter drafts done at by due date	Posts shortly after due date	Posts a couple of days into discussion	Posted, but too late for discussion to occur

Figure 1. Discussion grading rubric

The students in both the blended and online sections were graded based on all three categories. All students had the rubric prior to doing the assignment. The rubric scores ranged from 1 = poor to 4 = excellent for each category as shown in Figure 1. For grading purposes, these scores were converted to a 100% scale as shown in Figure 1.

3.3. Course activities

Two grades reflective of collaboration were included in this study. The first grade was a collaboration grade assigned by the instructor after tabulating peer and self-evaluations conducted after every collaborative class activity in all learning environments. On each evaluation, this grade could range from 5-20 points; the multiple evaluations were averaged converted to a grade out of 100. The evaluation areas included: *researches and gathers information*, *shares information*, *responsible and timely*, *participates in discussions*, and *shares workload*. The second collaboration grade was the final grade assigned by the instructor after evaluating a collaborative case study assignment that students worked on throughout the semester. The final project grade could range from 0-100.

Two summary grades were also included in this study. The first was a summary grade for all discussions in the class; they were graded by the professor and combined into an overall discussion grade that could range from 0-100. This discussion grade for the students in the online and blended sections was influenced by factors not included in the traditional class grade (timeliness of posts, number of responses to other discussion group members). These factors were viewed as inherent to the learning environment and included, although these same factors were not factors in the face-to-face discussion grade. The other summary grade was the final numerical grade for the entire course; this is the average of all the weighted individual grades.

3.4. End of semester survey

An instructor developed survey provided data on student opinions about the structure and design of the course, learning activities, and student and instructor interactions. The questions included the following:

1. What specific things do you think could be improved in the structure or design of the course and its learning activities?
2. How would you change the quality of and participation in course discussions/interactions?
3. What changes would you suggest to the pacing and sequencing of the content or activities in the course?
4. Do you have any suggestions how the course could be improved in terms of instructor interaction, participation, or management of the course?
5. What other suggestions, comments, or recommendations do you have about the course?

4. Procedure

Archival course data consisting of the end-of-semester surveys, student grades, and initial written responses to discussion prompts from all three sections of a graduate-level special education course were collected during each semester. Student data were coded and identifiers removed. Demographic information about students' backgrounds was gathered from student admission and graduation reviews and added to the coded files.

4.1. Discussion scoring with SOLO Taxonomy

The initial written responses to discussion prompts from two integrated discussions were coded for level of complexity using the SOLO Taxonomy. After establishing acceptable inter-rater reliability using student artifacts unrelated to the proposed study, two researchers previously trained to use this coding system conducted independent and blind coding of the initial written responses for the two integrated discussions for all students in the three sections. Each initial written response received one SOLO code ranging from 1 to 5 with 1 representing the lowest level of complexity (prestructural) and 5 representing the highest level of complexity (extended abstract).

Any disagreements in coding were reviewed and negotiated by the two coders; a third trained coder was to be consulted if disagreements could not be successfully negotiated. The third coder was not needed since all negotiated codes reached 100% agreement. Inter-rater percent of overall agreement before negotiations for Discussion 1 and Discussion 2 was 48% for each discussion and 100% after negotiations. While the percentage of agreement prior to negotiations was lower than anticipated, the majority of the negotiation discussions involved either evidence in the passages that was missed by one of the researchers or whether the response was a high or low quality within a specified level. For example, a low Level 5 (Relational) response may

begin to address issues related to the future, begin to generalize to another population, or provide solutions, but may not have provided the level of depth for a solid Level 5 response. A sample researcher negotiation comment for a low Level 5 response includes "does come up with a solution - not as in depth or creative but a solution". Comments such as this were included during the coding process. Including low and high codes for Levels 4 and 5 in particular may have improved inter-rater agreement on the SOLO coding; however, discussions regarding discrepancies were not time intensive indicating that agreements were generally easily and quickly negotiated.

4.2. Discussion scoring rubric

The same initial written responses from the two integrated discussions were re-rated for quality using only the quality indicator on the instructor-created grading rubric. After establishing acceptable inter-rater reliability using student artifacts unrelated to the proposed study, two researchers trained to use this rubric conducted independent and blind coding of each of the two integrated written responses for all students in the three sections. Each initial written response received one rating ranging from 1 to 4 with 1 representing poor quality and 4 representing excellent quality based on the first category on the rubric (Quality of Discussion Starter). Any disagreements in coding were reviewed and negotiated by the two coders; a third trained coder was to be consulted if disagreements could not be successfully negotiated. A third coder was not needed since all negotiated codes reached 100% agreement.

Inter-rater percent of overall agreement before negotiations for Discussion 1 and Discussion 2 was 63% and 65% respectively and 100% after negotiations. Although the percentage of agreement prior to negotiations was higher than the calculations for the SOLO, they were still lower than anticipated. The majority of the negotiation discussions involved either evidence in the passages that were missed by one of the researchers or a miscalculation of the number of sources cited. Unlike the SOLO coding, the number of sources cited had a significant impact on the scoring using the rubric. Discussions regarding discrepancies were not extensive indicating that agreements were generally easily and quickly negotiated.

4.3. Instructor assigned course grades

All instructor assigned course grades for the participants were gathered from the end of semester EXCEL grade spreadsheets and entered onto the coded data sheets for later entry analysis by SPSS. Grades compiled included discussion grades, collaboration grades, and the final course grade.

4.4. End of semester survey data

The end-of-semester survey designed by the instructor provided data on student opinions of improvements in the structure of the course and its learning activities; the quality and participation in course discussions and interactions; the course pacing and sequencing; and instructor interaction, participation, and course management. An analysis was conducted on each of the open-ended survey responses using the procedures as outlined by Strauss & Corbin (1990) including open coding, axial coding, and selective coding.

Open coding is the "process of breaking down, examining, comparing, conceptualizing, and categorizing data" (Strauss & Corbin, 1990, p. 61). The identified concept labels are derived by "taking apart an observation, a sentence, a paragraph, and giving each discrete incident, idea, or event a name, something that stands for or represents a phenomenon…[and in doing so will ask] What is this? What does it represent" (1990, p. 63). After the open-coding process, the labels were placed into categories or subcategories, also referred to as axial coding; during the axial coding process, the previously identified labels were carefully examined and grouped according to common themes or characteristics (Strauss & Corbin, 1990). Finally, the third stage is a process of selective coding that involved the identification of the core or "central category around which all other categories are integrated" (Strauss & Corbin, 1990, p. 116). For example, in the case of the survey data the central category may involve some component of the course that stands out as having had a unique impact on students based on the method of course delivery. It was at this time that the researcher used this qualitative data to 'tell a story' about student perceptions of the course supported by the categories identified within the overarching core or central category.

As only one researcher conducted the open, axial, and selective coding to "tell a story", a procedure described by Creswell (1998) was used to enhance the trustworthiness of qualitative findings – an external audit. Creswell explains that "external audits allow an external consultant, the auditor, to examine both the process and the product of the account, assessing their accuracy. The auditor had no connection to the study. In assessing the product, the auditor examined whether or not the findings interpretations and conclusions are supported by data" (Creswell, 1998, p. 203).

5. Results

Archival course data was utilized in this exploratory study and included student discussions, instructor-assigned course and assignment grades and end-of-semester surveys from all three sections of a graduate-level special education teacher preparation course. Analyses of the data are discussed in the sections that follow.

5.1. Discussion-related performance

To address the first research question, "Are there differences in level of understanding and performance among graduate students in a special education course delivered in three different learning environments?" a one-way multivariate analysis of variance (MANOVA) was conducted to determine the effect of the learning environment on the dependent construct of understanding. The fixed categorical independent variable was the learning environment with three levels (traditional, blended, and online). The continuous dependent variable was the construct 'level of understanding' and consisted of researcher ratings on two written discussion responses using the Structured of Observed Learning Outcomes (SOLO) Taxonomy (Biggs & Collis, 1982) that measured the complexity of responses; in addition, the researcher re-coded the two discussion responses using a portion of the instructor-designed discussion rubric related to the quality of the responses. Presented in Table 3 are the means and standard deviations of the researcher SOLO and rubric ratings for the two discussion responses. Higher means indicate more complex or higher quality responses.

When the descriptive statistics results for the researcher SOLO and revised Rubric scores on the two discussion responses were compared with the instructor-assigned grades for the same tasks, an interesting pattern emerged. Student performance in the traditional and blended learning environments was comparable; however, the scores for the students in the online section of the course were generally markedly lower than the other two sections of the course. In general, the students in the traditional course appeared to outperform the students in the blended and online sections of the course based on the mean scores for the researcher-assigned SOLO and Rubric scores. Generally, the SOLO score averages for each section fell within the Multi-Structural and Relational levels which appears to be in line with the existing literature, though it was anticipated that graduate students would score at the Extended Abstract level [13, 14]. Further analysis of the SOLO and researcher Rubric scores for the discussions provided a more in-depth review of the findings.

Table 3. Means and standard deviations for researcher rating of discussions by learning environment

	Traditional n=9	Online n=10	Blended n=20
	M (SD) Range	*M (SD) Range*	*M (SD) Range*
SOLO Discussion 1	3.78 (.83) 3-5	3.70 (1.16) 2-5	3.95 (1.05) 2-5
SOLO Discussion 2	4.22 (.83) 3-5	3.00 (.47) 2-4	3.70 (.86) 2-5
Rubric Discussion1	3.56 (.72) 2-4	2.70 (1.16) 1-4	2.50 (1.10) 1-4
Rubric Discussion 2	3.64 (.67) 2-4	2.10 (.31) 2-3	3.40 (.82) 2-4

Note: SOLO Taxonomy scores range from 1-5 with 5 representing the highest level. Rubric scores range from 1-4 with 4 representing the highest level.

The Barlett's Test of Sphericity was used to determine if a MANOVA was more appropriate than multiple univariate analyses (Tabachnick & Fidell, 2007). The Bartlett's Test examines correlations between the dependent construct and was determined to be significant, $\chi^2(4) = 48.39$, $p < .001$, indicating that it was appropriate to proceed with the MANOVA analysis. The application of Wilk's Lambda was used to evaluate multivariate significance to guarantee robustness of the results (Tabachnick & Fidell, 2007). The one-way MANOVA revealed significant differences among the three learning environments on the dependent construct (Wilks' $\Lambda = .40$, F (2, 36) = 4.69, $p < .001$). Post hoc univariate analyses for each dependent variable were examined as follow-up tests to the MANOVA and are presented in Table 4. Examination of the univariate analyses indicate two major contributors to the construct of understanding on both ratings of Discussion 2 responses (eg., Discussion 2 Rubric scores, F (2, 38) = 13.7, $p < .001$, $\eta^2 = .43$ and Discussion 2 SOLO scores, F (2, 38) = 5.9, $p < .01$, $\eta^2 = .24$). Variables reflective of the first discussion (Discussion 1 Rubric scores, F (2, 38) = 1.5, $p = .23$, $\eta^2 = .07$, and Discussion 1 SOLO scores, F (2, 38) = .22, $p = .80$, $\eta^2 = .01$) were not major contributors.

Scheffe post hoc comparisons were used to determine the location of the differences within learning environment for the Discussion 2 SOLO and rubric scores (see Table 5). There were no significant differences between the traditional and blended environments on both the Discussion 2 Rubric and SOLO scores or between the online and blended environments on Discussion 2 SOLO scores ($p > .01$). In contrast, the post hoc comparisons revealed significant differences for the online and traditional environments ($p < .001$) and online and blended environments ($p < .001$) on the Discussion 2 Rubric scores. In addition, there were significant differences

between the online and traditional environments ($p <$ .01) on the Discussion 2 SOLO scores.

Table 4. Summary of MANOVA and univariate analyses for environment by SOLO and rubric ratings for discussion 1 and discussion 2 responses

Source	*Multivariate* *df=4,36*	*Univariate* *df=2,36*	*Strength of Association* *Partial* η^2
Level Understand	4.69**		.59
SOLO Disc 1		.22	.01
SOLO Disc 2		5.97*	.25
Rubric Disc 1		1.50	.07
Rubric Disc 1		13.72**	.43

Note. Post-hoc univariate tests were conducted based on significant MANOVA statistics. *$p < .01$. **$p<.001$.

Fifty nine percent of the variance in the researcher ratings of students' level of understanding on Discussion 1 and Discussion 2 responses was attributed to the type of learning environment with statistically significant differences only found on Discussion 2 scores. If the learning environment were solely responsible for the differences in understanding, then it would be expected that statistically significant differences would have been found for both Discussion 1 and 2. There are a number of possibilities that might explain these differences; two possibilities include the format of the instruction and the learning environment.

Table 5. Summary of significant findings for discussion 2

	Significant Finding	
	Traditional	Online
Discussion 2 Rubric		
Traditional	-	Significance
Blended	No significance	Significance
Discussion 2 SOLO		
Traditional	-	Significance
Blended	No significance	No significance

Aspects of the format of instruction that could be attributed to the differences include the feedback received after Discussion 1, or variation in the content and prompts for Discussion 2 that may have been better structured to lead to higher levels of thinking. Biggs & Tang's *Principles of Constructive Alignment* focuses specifically on what should be done to advance the learning experience for students [15]. The authors stress that teaching and learning assessments need to be "designed and written with a view to the kind of knowledge and the level of understanding intended" [15]. Verbs such as theorize, reflect, generalize, and solve have a tendency to elicit higher levels of thought versus identify, define, and name which simply involve the process of recall and a lower level of thought [15].

The instructions for Discussion 1 asked students to consider whether *we should be moving in this direction* and *how can we prevent this in the future?* (course materials). These two elements from the discussion prompt would likely encourage students to move beyond the process of recall toward a deeper level of thinking including theorizing and inventing solutions as they look toward the future. In contrast, the instructions for Discussion 2 were not as explicit; students were asked to discuss how culture and families impacted the characters or the families. This could have lead to some theorizing but it appears to be eliciting factual information based on specific events, though the expectation of the researcher was that graduate students should not require that level of prompt. Additionally, the instructor grades and researcher ratings for both discussions were consistently lowest for students in the online course, suggesting that the learning environment may have been a key factor.

The discussions for the online and blended sections of the course were held completely online requiring students to visit and respond multiple times for each discussion in order to keep up with the 'conversations'. For both Discussion 1 and Discussion 2, the initial responses from the students in the traditional section of the course were generally longer, more detailed, and written more like a paper versus a response to begin a conversation with peers. The initial responses created by the students in the blended and online sections of the course were often more discussion-like in contrast to the responses made by the students in the traditional class; however, the students in the blended section had a tendency to create more detailed responses than the online students. Some students in the online and blended sections of the course did elaborate more in follow-up posts, but those responses were not included in the analyses for this study since there was no comparable artifact for the traditional section of the course. One possible explanation for better performance by the students in the blended section is that these students benefited from the face-to-face interactions where further clarification of the instructions was possible. There is some evidence from the student surveys that suggest this may have been the case. For example, one student mentioned there was "too much time spent on online housekeeping". While students in the online section of the course had opportunities to pose similar questions, not everyone in the class would necessarily see or have access to that information since all communications were in a written format. This is a potential problem that is unique to online learning environments.

Additionally, there is evidence from the student surveys that the students in the online section of the course struggled a great deal with the discussions stating that "online discussions were confusing and time consuming" and "multiple postings were intimidating and frustrating" and "asynchronous

postings were frustrating and time consuming". In contrast, the comments from the students in the blended section suggested they were generally satisfied with the online discussions. For example, two respondents stated that they "enjoyed online discussions – groups talked and elaborated more" and "discussion participation and quality [was] very good". The comments from students in the online and blended learning environments support the statistical evidence that suggests the learning environment may be responsible for the differences in student understanding and performance related to the discussions.

5.2. Instructor-assigned grades

To answer the second research question, "Are there differences in course grades among graduate students in a special education course delivered in three different learning environments", descriptive statistics and one-way analyses of variance (ANOVA) were utilized to examine differences in grades assigned by the instructor. Presented in Table 6 are the means, standard deviations, and range of scores for Discussion 1, Discussion 2, Final Discussion Grade, Final Collaboration Grade, Final Group Project Grade, and Final Course Grade.

Although there appears to be some variability in scores within the online course and between the online course and the blended and traditional course for Discussion 2 and the Final Discussion Grade, examination of Table 6 indicates an overall lack of variability in scores for the discussion-related measures in the traditional course (Discussion 1, Discussion 2 and Final Discussion Grade), indicating that analysis of variance was not appropriate for these variables.

Table 6. Means and standard deviations for instructor assigned discussion grades by learning environment

	Traditional (n=11)	Online (n=11)	Blended (n=20)
Instructor Grades	*M (SD) Range*	*M (SD) Range*	*M (SD) Range*
Discussion 1	100.00 (.00) 100	67.73 (26.17) 11-100	100.00 (.00) 100.0
Discussion 2	100.00 (.00) 100	74.91 (16.34) 33-100	95.10 (5.66) 83-100
Final Disc Grade	100.00 (.00) 100	85.09 (10.97) 62- 97	96.20 (6.34) 72-100
Collaboration Grade	97.82 (2.04) 95-100	89.55 (8.50) 75-100	98.90 (2.33) 90-100
Group Project	93.27 (2.64) 90- 98	93.64 (1.69) 92-96	93.64 (1.69) 87- 98
Final Course Grade	96.82 (1.25) 94-98	90.73 (4.88) 82-97	95.10 (2.24) 90-98

A pattern similar to the one found for the discussion responses also emerged for the instructor-assigned group Collaboration Grade and the Final Course Grade, but interestingly, not for the Group Project Grade where the grades were similar across all three learning environments. One possible explanation for the similarities in the Group Project Grade is that the work of multiple students influenced grades received on the project, whereas the Collaboration Grade was based on individual performance within the group. Of course, the Final Course Grade is directly influenced by all other grades received in the course. Of interest is the variability of scores within the online environment on Collaboration; the range of scores was 75 to 100 whereas the lowest score in the other two sections was 90. One possible explanation for the differences in scores is that the instructor could see the level of participation in the online section of the course; therefore, it would be more difficult for students in the online section of the course to inflate the collaboration scores on the activities for the group project.

While assimilation of the course content was necessary (primarily materials from readings and textbook information), the group collaboration process involved negotiation and creation of an end product. It required more sustained effort and continuous problem solving on the part of the group members with multiple iterations of the product before it was complete. The group project was more task-oriented whereas the group discussions are more conceptually-based. The results of the analyses suggest that collaboration in a fully online environment tends to be more difficult than in traditional and blended learning environments.

The Collaboration Grade related to the Group Project activities; some students in the online section of the course reported having difficulties with collaboration and suggested "more individual versus group projects", "reducing or combining some activities", and one student said s/he "would recommend others take [the course] on campus". This is further supported by similar comments made by some students in the blended section of the course which included "more in-class time with group", "less online discussions", and "difficult to keep track of so many components", suggesting that online collaboration might be more problematic than in the traditional face-to-face learning environments. Of interest is that the students in the blended section of the course were also afforded the opportunities to collaborate during in-class sessions, yet some still described online work as problematic. However, students in the blended section did not meet on campus every week since part of the 'class time' was spent online. The issues raised by the respondents in the traditional section of the class related more to the amount of work and the inclusion of more details and examples for the project. Fewer issues were mentioned regarding collaboration likely due to the fact that students met each week and had the

opportunity to collaborate in class on a regular basis. To increase student satisfaction and improve performance, online collaborative activities likely need to be restructured, perhaps with opportunities for synchronous interactions through chat rooms and video conferencing.

5.3. Student understanding, grades and demographics

The researcher assigned SOLO and Rubric scores for the two discussions were significantly correlated for both the online and blended learning environments, meaning when the score on one went up, the score on the other did as well. The correlations in both sections and high power results from the blended section might relate to the reliability of the instruments, indicating that the SOLO Taxonomy was well-aligned with the portion of the instructor-designed discussion rubric related to the quality of the response. However, this does not explain why there was no significant correlation for the same measures used in the traditional learning environment. This may be the result of the overall higher scores with less variance in the traditional environment than the online and blended environments. For example, the range in scores for the traditional environment on Discussion 1 for the SOLO was 3-5, whereas the range in the other two environments was 2-5. A further detailed analysis of the scores by individual might help to explain the lack of correlation SOLO and Rubric scores for the traditional environment.

In both the traditional and blended environments, the Group Project and Final Course grade were positively correlated which would be expected as the Group Project is an integral part of the course. However, in the online environment, it was the Collaboration Grade that was correlated with the Final Course grade. One possible explanation for this difference is that student's collaboration assessments in the online environment were submitted in the private assignment section online; this may have resulted in more honest peer assessments. Additionally, the instructor was better able to monitor participation in collaborative activities in the text-based, asynchronous environment online. However, it does not negate the fact that collaboration is more difficult in online learning environments, perhaps primarily due to the time waiting for others to respond.

6. Conclusion

Unlike the existing literature on online instruction in the field of special education, the findings in the exploratory study indicate that student performance was notably different in the traditional versus online sections of this course with students in the traditional course outperforming the students in the online version. However, the results for the blended and traditional learning environment comparisons indicate that students in both sections performed equally as well which supports other similar studies [16, 17]. Some researchers suggest that students in blended learning environments may outperform students in traditional settings; however, that did not hold true for this study which suggests student performance in the traditional and blended learning environments were very similar and student performance was the lowest in the online learning environment on the majority of the measures. The student survey data in this study suggest that the structure and design of the course might be a serious consideration and provide a possible explanation for the differences.

An important lesson learned in this study is that while the same content and activities can be utilized in the same course taught in three different learning environments, the content and its presentation, teaching methods and requirements for demonstrating understanding may need to be adapted based on the learning environment. Discussions, although not easily monitored by the instructor, are less time-consuming for students in face-to-face settings than they are for students in online classes. Online discussions are typically scrutinized more by instructors and often include penalties for timeliness and depth of responses. Perhaps the type and amount of collaborative activities and discussions should be considered for this and similar online courses along with grading criteria modifications.

7. Implications for practice

Hammond conducted a review of curriculum design, assumptions about teaching and learning, and conditions for using online discussions in a sample of case studies that used asynchronous online discussions in higher education [18]. The author including issues related to course design, what instructors should do, what learners should know, and what technology should offer. Many of the positive attributes described by Hammond were already a strong part of the course in the current study including, providing formative peer assessment, summative assessment of group products, explicit learning outcomes for group work, summative assessment of process and credit for participation, requirements for minimum level of participation, setting explicit tasks, review of group work process, and rotating roles within groups [18]. However, two issues raised by Hammond that may have been missing key elements in the online course design were the adjustment of the workload to allow time for discussions and collaboration and student learning preferences. Fink also proposed factors and

processes required to design significant learning experiences for students [19].

The first part of the recommended process by Fink is to conduct an in-depth analysis of situational factors to define constraints and opportunities for learning including, for example, the content to be taught and the characteristics of the learners and instructor [19]. The second step in the design process involves the development of learning goals that focus on significant learning, not just 'understand-and-remember' type of activities; these must include educative feedback and assessment with opportunities for students' self- assessment using clear criteria and standards. Fink recommends that teaching and learning activities engage students in experiential and reflective learning while integrating and aligning all of the major components of instruction, including situational factors, goals, feedback and assessment. The situational factors, in part, refer to the characteristics of the students, when, where, and how the course will be delivered, and the pedagogical challenges [19].

Workload and time factors were frequently cited issues of concern for the students in both the blended and online sections of the course but were most prevalent for the students in the online course. If online discussions are going to be scrutinized and penalized for quality of content, and frequency and timeliness of posts, students will need to spend more time creating thoughtful responses and visiting the discussions. This may require the reduction or reorganization of the workload. This is in line with the research of Chickering & Gamson and Chickering & Ehrmann who emphasized time on task by allowing adequate time for teaching and learning, as well as communicating high expectations, and respecting diverse talents by allowing opportunities for students to show their talents and learn in ways that work for them [20, 21]. This can only be accomplished through thoughtful planning of course structure, design and expectations.

Lastly, workload, time, student experience, and student learning preferences are some of the factors that were not explored in this study nor were they considerations when students were assigned to the various sections of the course. These important factors warrant further evaluation during the planning and implementation stages of the course design processes.

8. Strengths and limitations of the study

The strengths of this exploratory study included consistency across all three sections of the course (traditional, online, and blended) in the instructor, course materials, activities and assignments. In addition, all participants were in the final semester of the same special education teacher education program. Analyses revealed that the students in all three sections were well-matched in terms of their age, undergraduate GPA, and writing ability. During the literature review process, no studies were found that matched these criteria; however, there are some limitations that warrant discussion.

First, this study utilized archival data that were collected and analyzed after the courses had ended. This fact limited the researchers' ability to control variables and ensure the collection of all related data. Additionally, student interviews and classroom observations would have strengthened the study. Another limitation of the study was the small and unequal sample sizes that may have had an impact on the results. Given these limitations, the results of this exploratory study should be interpreted with caution.

9. Suggestions for future research

As the discussion threads for both the online and blended courses have been maintained, future analyses will evaluate these data to further examine the responses as the students continued the discussions. There are also a number of issues that could be addressed to strengthen future studies, including controlling for equal and larger sample sizes if possible and the collection of data while the courses are in progress. Additionally, classroom observations and video recordings of in-class discussions and instructor and student interactions, access to online and email communications between the instructor and students regarding questions and concerns on coursework, access to the online discussions, and pre/post assessments of knowledge or other relevant information would further strengthen this and future studies. It is recommended that future studies continue to investigate multiple learning environments, and examine in more detail students' characteristics and past experiences, particularly related to distance learning. These studies could involve varying course designs and assignments to further explore the impact of workload and time on student opinion and performance.

10. References

[1] United States Department of Education (2009), Office of Planning, Evaluation, and Policy Development. Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies, Washington, D.C.

[2] Caywood, K. & Duckett, J. (2003). Online vs. on-campus learning in teacher education. Teacher Education and Special Education, 26(2), 98-105.

[3] O'Neal, K., Jones, W.P., Miller, S.P., Campbell, P., & Pierce, T. (2007). Comparing web-based to traditional instruction for teaching special education content. Teacher Education and Special Education, 30(1), 34-41.

[4] Skylar, A.A., Higgins, K., Boone, R., Jones, P., Pierce, & Gelfer, J. (2005). Distance education: An exploration of alternative methods and types of instructional media in teacher education. Journal of Special Education Technology, 20(3), 25-33.

[5] Steinweg, S.B., Davis, M.L., & Thomson, W.S. (2005). A comparison of traditional and online instruction in an introduction to special education course. Teacher Education and Special Education, 28(1), 62-73.

[6] Bullock, Gable, & Mohr (2008). Technology-mediated instruction in distance education and teacher preparation in special education. Teacher Education and Special Education, 31(4), 229-242.

[7] Creswell, J.W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches. Los Angeles, CA: Sage Publication.

[8] Allan, J. & Street, M. (2007). The quest for deeper learning: An investigation into the impact of a knowledge-pooling WebQuest in primary initial teacher training. British Journal of Educational Technology, 38(6), 1102-1112.

[9] Biggs, J.B. (1979). Individual differences in study processes and the quality of learning outcomes. Higher Education, 8, 381-394.

[10] Biggs, J.B. (1989). Approaches to the enhancement of tertiary teaching. Higher Education Research and Development 8, 7-25.

[11] Delfino, M. & Persico, D. (2007). Online or face-to-face: Experimenting with different techniques in teacher training. Journal of Computed Assisted Learning, 23, 351-365.

[12] Kanuka, H. (2005). An exploration into facilitating higher levels of learning in a test-based internet learning environment using diverse instructional strategies. Journal of Computer-Mediated Communication, 10(3), article 8.

[13] Boulton-Lewis, G.M., Wilss, L.,& Mutch, S. (1996). Teachers as adult learners: Their knowledge of their own learning and implications for teaching. Higher Education, 32, 89-106.

[14] Thomas, M.J.W. (2002). Learning within incoherent structures: The space of online discussion forums. Journal of Computer Assisted Learning, 18, 351-366.

[15] Biggs, J.B. & Tang. (2007). Teaching for Quality Learning at University. Berkshire & New York: Society for Research into Higher Education & Open University Press.

[16] Dowling, C., Godfrey, J.M., & Gyles, N. (2003). Do hybrid flexible delivery teaching methods improve accounting students' learning outcomes? Accounting Education, 12(4), 373-391.

[17] Larson, D.K. & Sung, C.H. (2009). Comparing student performance: Online versus blended versus face-to-face. Journal of Asynchronous Learning Networks, 12(1), 31-42.

[18] Hammond, M. (2005). A review of recent papers on online discussion in teaching and learning in higher education. Journal of Asynchronous Learning Networks, 9(3). http://www.sloan-c.org/publications/jaln/v9n3/v9n3_hammond.asp

[19] Fink, L.D. (2003). Creating significant learning experiences. San Francisco, CA: Jossey-Bass

[20] Chickering, A.W. & Gamson, Z.F. (1987). Seven principles for good practice in undergraduate education. AAHE Bulletin, March, 3-7.

[21] Chickering, A.W. & Ehrmann, S.C. (1996). Implementing the seven principles: Technology as a Lever. AAHE Bulletin, October, 3-6.

