SUPPORTING NOVICE LEARNING THROUGH FOCUSED PROCESSING OF WORKED EXAMPLES AND EXPLANATIONS

By

Katherine L. McEldoon

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

In partial fulfillment of the requirements

For the degree of

DOCTOR OF PHILOSOPHY

In

Psychology

May, 2014

Nashville, TN

Approved:

Professor Bethany Rittle-Johnson

Professor Bruce McCandliss

Professor Georgene Troseth

Professor Leona Schauble

ACKNOWLEDGEMENTS

Primary thanks undoubtedly go to Bethany Rittle-Johnson. I could not have had a better mentor to guide, model, and support me through this process. I owe this work and who I am as a researcher to her. Bethany's incredibly sharp mind coupled with her supportive mentoring style has been instrumental to both my professional and personal development. She sets the bar extraordinarily high as a role model, but it is a bar I strive to come at least a little close to. Thank you for pushing me, encouraging me, and believing in me.

Many thanks go to my committee members as well. Their insightful comments and valuable feedback have greatly improved this work. To Georgene Troseth, thank you for showing me that being a researcher is a process, not an end product. Your insightful guidance has made a large impact on how I approach my work and my life. To Bruce McCandliss, thank for you sharing your never-ending excitement for science, and that there's always another fascinating question around the corner. To Leona Schauble, thank you for challenging me to think from another perspective, and for pushing me to be a refined and careful thinker.

I am forever grateful to all my lab members I've had the honor to work with, whose support and camaraderie helped me every step of the way. Percival Matthews, thank you for always letting me push your big brother button, and for either making me laugh or annoyed, as you laugh back at me. Kelley 'you're the best!' Durkin, it has been beyond a pleasure to work with you. Thanks for always being one step ahead of me, and for always helping me figure out what I need to do next. Emily Fyfe, your hard work and focused nature are an inspiration; thank you for sharing that with me. I would also like to thank postdocs Marci DeCaro and Michael Miller, whose guidance and support have been invaluable. Thank you for being such great role models. Thanks to lab managers Laura McLean and Abbey Loehr, without whose on-the-ball organization and support, this work would not be possible. Finally, many thanks to the undergraduate members of our lab team; their countless hours of work and ever-vigilant attention to detail have been instrumental to my success.

I would like to acknowledge my funding agency. The research reported here was funded by the Institute of Education Sciences, U.S. Department of Education, Grant #R305B040110. The views expressed in this paper are mine and not theirs. In addition, thank you to all of the schools, teachers, and students who participated in my research.

I'd like to give a huge thanks to all of my wonderful, inspiring, fun, and incredibly supportive friends. I am beyond grateful for all the love and laughter you've provided me over the last few years. Thank you for always being willing to listen, to celebrate with my victories, and commiserate with my losses. Among many others, a special thanks to Amber Chatelain, Stacy Davidson, Kyle Gorman, and Kyle Blauw. You all deserve a deep heartfelt thanks. To Lance Conzett, thank you for your love, support, patience, and laughter. You make my every day better. You all have made this journey extraordinarily exciting and worthwhile.

Finally, I would like to thank my family. To my sister Allison, thanks for taking a chance and moving to Nashville with me. Thank you for putting up with me being a stressed out grad student, I know I wasn't always easy. Despite it all I love you and I'm so happy we've been able to live our lives together. And to my parents Joan and Joel, thank you for always trusting that I was making the right decisions, for providing me with such a great foundation, for your continual guidance and support, and for loving me no matter what. I am beyond lucky to have such a great family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

When learning a new domain, the learner often begins as a novice. The goal of learning is to transition out of the novice state and into a more knowledgeable and competent state. Within problem solving domains, one well-established way of fostering learning in novices is through the use of worked examples (e.g. Atkinson, Derry, Renkl, & Wortham, 2000; Sweller, 2006). However, simply viewing worked examples is not enough to foster meaningful learning. Contemporary educational theory places a high value on the learner engaging in active cognitive processing during learning (Mayer, 2009). However, *what* the learner should focus their processing on to maximize learning outcomes is currently underspecified. This project compares two levels of processing demands when studying worked examples to support the transition from novice to more advanced learner within introductory statistics.

Worked Examples Support Learning in Novices

One proven way to support novice learning in problem-solving domains is through the use of worked examples. Worked examples are instructional devices that include a problem statement and an expert procedure for solving the problem. They have often been used within domains such as mathematics, physics, and computer programming (Berthold & Renkl, 2009; Catrambone & Yuasa, 2006; Hilbert, Renkl, Kessler, & Reiss, 2008; Moreno, 2006). Typically, there are three components: the problem statement, the solution steps undertaken, and the solution (Atkinson et al., 2000; Renkl, Stark, Gruber, & Mandl, 1998). When initially learning in a new domain, studying a worked example can be an effective way to help novices learn to solve problems (Sweller, 2006). Nearly 20 years of research has consistently shown that worked examples not only represent a highly-valued source of instruction by learners (e.g., LeFevre &

Dixon, 1986; Pirolli & Anderson, 1985; Recker & Pirolli, 1995) but an effective one as well (for an overview see Atkinson et al., 2000). Worked examples are particularly effective for novice learners who have low prior knowledge of the domain (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 2000; Kalyuga, Chandler, Tuovinen, & Sweller, 2001). Compared to solving problems on one's own, studying worked examples leads to greater learning among novices in many domains: in third grade arithmetic (Mwangi & Sweller, 1998), in high school and university algebra (Carroll, 1994; Sweller & Cooper, 1985); high school geometry (Paas & van Merriënboer, 1994; Tarmizi & Sweller, 1988); university statistics word problems (Quilici & Mayer, 1996); and university-level engineering courses, such as series and parallel electrical circuit analysis (Reisslein, Atkinson, Seeling, & Reisslein, 2006); geometric optics and kinematics (Ward & Sweller, 1990); and electrical circuits troubleshooting (van Gog, Paas, & van Merriënboer, 2006). Overall, learning from worked examples is more effective for problem-solving skill acquisition by novices than simply engaging in problem-solving alone. This benefit has been formalized by Sweller, Merrienboer, & Paas (1998) as *the worked example effect.*

Learning from worked examples has been shown to be especially important and effective during initial skill acquisition within well structured domains (e.g. physics, mathematics, programming) (Atkinson et al., 2000; Moreno, 2006). Worked examples can be useful because novice learners often attempt to solve problems by analogy. They use problems they already know how to solve as examples, find relationships and similarities between the known and new problems, and apply problem-solving strategies from the known examples to solve new problems. Overall, worked examples provide students with an example to use as an analog when solving a new problem type. This then frees working-memory and attentional resources to process structural aspects of the problem instead (Van Lehn, 1998). Carefully designed worked examples can help learners go from a novice state, support the acquisition of the structural foundations and problem solving procedures of the domain, and bootstrap the learner into a more advanced knowledge state (Anderson et al., 1997).

Focused Processing of Worked Examples

Although studying worked examples is beneficial for learning (Atkinson et al., 2000; Sweller, 2006), these benefits are even greater when there are built-in requirements for direct student input and engagement, or focused processing (Atkinson & Renkl, 2007). Researchers who advocate for focused processing assume that knowledge cannot be imparted on learners but instead must be actively constructed via information processing in working memory (Berthold & Renkl, 2010; Robins & Mayer, 1993). If they are not explicitly instructed to do so, some novice learners will engage with examples in a passive or superficial manner (Renkl, 1997). It is not common for novices to spontaneously engage in focused cognitive processing of examples, such as engaging in elaboration or comparison (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Gerjets, Scheiter, & Catrambone, 2006; Renkl, 1997). They may not be able to identify the most important information in the example and instead attend to irrelevant features (Ross, 1989). Novice learners also may succumb to an illusion of understanding, falsely believing that they understand the example (Renkl, 1999). For these reasons, novices benefit from scaffolds that support focused processing of worked examples.

Several studies have explored ways to scaffold focused processing when learning from worked examples (Atkinson et al., 2000; Moreno, 2006; Paas & van Gog, 2006; Renkl, 2002). I define focused processing as instructional scaffolds that focuses the learners' attention and encourages them to process relevant information. For example, prompting learners to generate explanations (i.e. self-explain) while studying worked examples facilitates greater learning of a computer programming language than studying the same examples with the answers to the explanation prompts provided (Catrambone & Yuasa, 2006). Other focused processing scaffolds include finishing incomplete worked examples with missing solution steps (Reisslein et al., 2006; Stark, 1999), labeling problem solving step subgoals (Catrambone, 1996, 1998), and explicitly linking and mapping different representations within the example (such as labeling a geometry diagram with the related algebraic formula) (Berthold & Renkl, 2009; Tarmizi & Sweller, 1988; Ward & Sweller, 1990; see also: Catrambone & Yuasa, 2006; Große & Renkl, 2006; van Gog et al., 2006). Prompting students to engage with and process relevant features of worked examples is a key facet of the worked example effect.

In an effort to highlight *what* relevant aspects the learners should be encouraged to process, Renkl & Atkinson (2007) proposed a *focused processing stance* that specifies that learners should focus their processing on the *central concepts to be learned* (e.g., mathematical theorems, physics laws) (Berthold & Renkl, 2010; Wittwer & Renkl, 2008). For example, learners should focus their processing on understanding geometric principles, not on identifying which phase of the proving process the example is in (Hilbert et al., 2008).

When prompts focus learners' processing on the central concepts, learning outcomes are improved. For example, when high school students learning about probability theory completed focused processing prompts in the form of "why" questions about *concepts* within provided instructional explanations, conceptual knowledge and transfer ability was greater at posttest relative to students learning without the focused processing prompts (Berthold & Renkl, 2008).

However, this line of research has revealed limitations to focusing processing on concepts alone. Berthold, Röder, Knörzer, Kessler, & Renkl (2011) had novice university tax law students work through an e-learning module with focused processing prompts on underlying concepts (i.e. prompts to generate conceptually-oriented explanation) and a control group that did not have these prompts. Inclusion of focused processing prompts lead to a positive effect for conceptual knowledge, as hypothesized. Importantly, there was a simultaneous *negative* effect for knowledge of problem solving *procedures*. The students who did *not* receive the focused processing prompts had significantly higher *procedural knowledge* scores at posttest. This same pattern of results was also found in a similar study by Berthold & Renkl (2009). Focused processing of concepts, though beneficial for learning these concepts, actually detracted from learning problem solving procedures. This is problematic, as learning both concepts *and* procedures are necessary for complete understanding. This highlights an important limitation of Renkl & Atkinson's (2007) focused processing stance.

Focused Processing of Concepts and Procedures

The goal of specifying what learners should focus their processing on is important. However, the focused processing stance of Renkl & Atkinson (2007) needs to be refined. I propose to widen the scope of critical features of the domain beyond concepts to combat the negative effects on procedural knowledge. In particular, I propose that processing should also be focused on assigning meaning to values within problem solving procedures.

Focusing attention on the meaning behind procedures entails understanding the *goals* of the steps in the procedure and how the specific *operators* (e.g. addition or subtraction) within the procedure accomplish these goals. This combination between the goals and the operators that are necessary to accomplish these goals has been formalized as *goal-operator combinations*. Understanding the link between the goal and the operators is key. Goal-operator combinations are "a way by which a learner can assign meaning to operators by identifying the subgoals achieved by those operators" (Renkl, 2011). For example, in a probability problem, the goal-operator elaboration might be: By subtracting (*an operator*) the probability of red items from 1, we get the probability of non-red items (*the goal*). Considering goal-operator combinations fosters the representation of goals (and sub-goals) to be achieved, including how the steps and operations within the procedure achieve these goals. Several studies suggest that elaborations on goal-operator combinations in worked examples foster transfer to novel problems (Catrambone, 1996; Chi et al., 1989; Conati & Van Lehn, 2000; Renkl et al., 1998; Renkl, 1997). For example, elaborations on goal-operator combinations during learning (e.g. "Through this multiplication we get the probability of tiles with color and form faults") was positively correlated with post-test performance, particularly for transfer (Renkl, 1997).

One way to foster processing of goal-operator combinations is to require leaners to determine missing problem-solving steps within incomplete worked examples. With this requirement, learners must reflect on what the goal of the problem-solving step is, and how the values and operators accomplish the goal. Indeed, having learners to fill in missing problem-solving steps improves learning and transfer relative to studying complete worked examples (Reisslein et al., 2006; Stark, 1999).

Unfortunately, novices often fail to attend to goal-operator combinations when studying worked examples. They often do not gain a deep understanding of how the solution steps relate to and achieve the problem solving goals (Chi et al., 1989; Renkl, 1997). Thus, prompting novices to focus on goal-operator combinations should foster deeper knowledge of procedures and how they link to the concepts of the domain.

To address the limitations present in the focused processing stance of Renkl & Atkinson (2007), I propose a *Modified Focused Processing Stance*. To maximize learning outcomes, learners should be scaffolded to process *both* the primary concepts to be learned, as well as goal-operator combinations within problem solving procedures. Processing that focuses on concepts involves recalling relevant concepts from the text and/or prior knowledge, and connecting them to the current material. Processing that focuses on goal-operator combinations involves assigning meaning to problem-solving steps by identifying how the operators achieve the relevant goals

(Catrambone & Yuasa, 2006; Renkl, 1997). The *quality* of focused processing a learning activity elicits can vary as a function how much or how deeply a learner engages with the concepts and goal-operator combinations.

If focused processing is an effective conceptualization, then instructional materials that elicit more or less of it (i.e. manipulate the quality of focused processing) should result in correspondingly higher and lower quality learning outcomes. Overall, this *Modified Focused Processing Stance* contains more breadth than either a focus on concepts or goal-operator combinations alone. It is hypothesized to overcome the limitations of the focused processing stance of Renkl & Atkinson (2007).

The current study will investigate two different types of focused processing modifications, each designed to elicit higher and lower quality focused processing. The study focuses on improving worked examples for helping novices learn introductory statistics. However, this framework could be applied more widely to other instructional domains.

Quality of focused processing could be varied in many ways. Two considerations are whether or not learners *generate* information, as opposed to reading it, and if the generated information requires inferences beyond the given information or is simply a summary without inference. Do learners generate explanations of underlying concepts or read explanations provided by others? Do learners generate missing information that requires an understanding of goal-operator combinations, that are central to understanding; or do they generate more shallow missing information that can be discerned locally, without assigning meaning or linking to goals and concepts? The higher quality information the learner generates, the greater his/her focused processing. This study uses two methods to varying the quality of focused processing. Each method is considered in turn.

Method 1: Reading Instructional Explanations With or Without Prior Self-Explanation

Including or inducing explanations of underlying concepts is one effective way to improve learning from worked examples (see Atkinson & Renkl, 2007 for a review; Berthold & Renkl, 2010; Wittwer & Renkl, 2010). Indeed, explanations are considered a critical feature of effective learning from worked examples (Atkinson & Renkl, 2007; Renkl, 2011).

Since worked examples explicitly lay out procedural problem solving steps, they are at times augmented with instructional explanations that justify the steps or provide conceptual information supporting those steps (Atkinson et al., 2000; Bielaczyc, Pirolli, & Brown, 1995; Paas & van Gog, 2006). Instructional explanations can be thought of as an attempt to give answers to questions that are implicitly or explicitly posed by learners or teachers, and are designed for the purpose of teaching (Duffy, Roehler, Meloth, & Vavrus, 1986; Leinhardt & Steele, 2005; Leinhardt, 2001; Treagust & Harrison, 1999).

Indeed, providing novices with explanatory information might be necessary for high-quality learning. Novices have insufficient prior knowledge to gain a high quality understanding from worked examples alone (Kirschner, Sweller, & Clark, 2006; Renkl, 2002; Wittwer & Renkl, 2008). Instructional explanations may help contextualize the problem solving steps into the larger learning domain and enhance understanding of solution procedures. That is, they enhance understanding of why solution steps are effective and/or of when they should be applied. Several studies have shown that the integration of instructional explanations into worked examples benefits learning compared to worked examples without instructional explanations (Catrambone, 1998; Gerjets et al., 2006; Schworm & Renkl, 2006; Stark, Mandl, Gruber, & Renkl, 2002). In a recent meta-analysis, providing instructional explanations in addition to worked examples is particularly effective for supporting knowledge of concepts compared to worked examples alone (Cohen's $d = 0.36$; Wittwer & Renkl, 2010).

However, instructional explanations can be ineffective when focused processing of them is not supported. Processing the instructional explanation seems to be key. For example, solving a follow-up question using the information in an instructional explanation was the strongest predictor of learning outcomes in a study on group mathematics learning in an elementary classroom (Webb & Farivar, 1999). Several studies have found instructional explanations to have no positive effects on learning outcomes (e.g. Chi, 2000; GroBe & Renkl, 2006; Hausmann & VanLehn, 2007; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003; Webb 1989). In a review of 21 studies of instructional explanations embedded within worked examples, Wittwer & Renkl (2010) noted large variability in whether the provision of instructional explanations aided learning across studies. There are several proposed reasons they may sometime be ineffective. One reason may be because novice learners often read them in a passive way, resulting in superficial processing (Berthold & Renkl, 2010). Instructional explanations may even lead to inhibition of other focused processing activities. For example, Schworm & Renkl (2006) found that the provision of instructional explanations decreased the amount of self-explanation activity (i.e., generating explanations to oneself in an attempt to make sense of new and known to be correct information (Chi, 2000)). Further, instructional explanations may cause an illusion of understanding, in which learners falsely believe they understand the material better than they actually do (Atkinson et al., 2000). This may reflect the "double curse of incompetence", where novice learners not only lack the to-be-learned information, but also lacks the metacognitive awareness of this deficiency (Dunning, Johnson, Ehrlinger, & Kruger, 2003). Wittwer & Renkl (2008) emphasized that instructional explanations should not replace learners' active or focused processing activities.

Instead of reading provided instructional explanations, learners can be prompted to generate their own explanations. This process is called self-explanation, and it often increases learning and transfer relative to not generating explanations (Chi, Leeuw, Chiu, & Lavancher, 1994; Fonseca & Chi, 2011; Rittle-Johnson, 2006). Self-explanation has been demonstrated as an effective learning technique in young children playing checkers (Calin-Jageman & Ratner, 2005), school-aged children learning mathematics (Matthews & Rittle-Johnson, 2009; McEldoon, Durkin, & Rittle-Johnson, 2013; Siegler & Lin, 2010; Wong, Lawson, & Keeves, 2002), high school and university students learning physics and probability (Große & Renkl, 2004; Hilbert et al., 2008), and student teachers learning how to design effective lessons (Große & Renkl, 2006).

Self-explanations vary in quality. Explanations that contain inferences about concepts, goal-operator combinations, or that anticipate upcoming problem-solving steps were related to greater learning outcomes (e.g. Chi et al. 1989; Renkl, 1997). However, generating such high-quality explanations can be difficult during initial skill acquisition. Novice learners may not have sufficient prior knowledge to produce these types of explanations. Often, novice learners produce vague or incorrect self-explanations, and this can impede or even harm learning (Berthold & Renkl, 2009; Conati & Van Lehn, 1999; Renkl et al., 1998). For example, when fourth-graders engaged in the challenging task of identifying causal effects by investigating a database, prompts to self-explain were actually detrimental to casual inference performance (Kuhn & Katz, 2009). This result is likely due to the students having insufficient prior knowledge to make productive use of the prompts to self-explain.

In sum, both instructional- and self-explanations both have their benefits and drawbacks. *Combining* self- and instructional-explanations may be one way to harness the benefits of both and promote more focused processing of worked examples.

Indeed, engagement with instructional explanations increases when used in combination with self-explanation prompts. For example, in a study on probability learning in high school students, worked examples and instructional explanations were provided with and without additional self-explanation prompts. Answering the self-explanation prompts required integration of the instructional explanations with the learners' prior knowledge and the learning environment. The combination of instructional- and self-explanations resulted in an increased discussion of domain principles and conceptual knowledge relative to optional note-taking and to no self-explanation prompts (Berthold & Renkl, 2010). If instructional explanations are used in tandem with self-explanations to increase focused processing by learner, they can be more effective as an instructional tool than unguided note-taking (Berthold & Renkl, 2010; Wittwer & Renkl, 2008). The roles can be reversed, and instructional explanations can augment self-explanations as well. In a study with high school students learning with a geometry cognitive tutor, if a student's self-explanation was incorrect, they were provided with instructional explanations, or hints, that became successively more directive if the student continued to struggle. This combination of self- and instructional explanations was more effective for correct problem solving, reasoning, and judgments about problem types than no explanation prompts at all (Aleven & Koedinger, 2002).

This study contrasts instructional explanations with and without prior self-explanation prompts. Is it best to provide a novice with instructional explanations immediately? Or is it best to prompt them to self-explain first, even though they may generate poor quality explanations, and *then* be presented with the instructional explanation so that they can revise their initial ideas? The learners in both cases ultimately receive the same instructional explanation. However, the second scenario entails a higher quality of focused processing, as the student *generates* an explanation. When students are asked to generate an explanation, they activate their prior knowledge and build an initial problem representation. When they are *then* given the instructional explanation and asked to self-correct their explanations, they are able to update their representation and correct their misconceptions. Overall, having learners generate their own explanations (i.e., prompts to self-explain) before receiving an instructional explanation should increase the quality of focused processing and subsequent learning than receiving instructional explanations immediately.

Method 2: Incomplete Worked Examples with Gaps in Goal-Operator Combinations vs. Computations

In addition to focusing on underlying concepts, novices need to focus on important aspects of the problem solving procedure, such as goal-operator combinations. One way to focus attention on goal-operator combinations is to omit key values from the worked example, and require the learner to fill them in (e.g., incomplete worked examples) (Paas, 1992; Stark, 1999; van Merriënboer & De Croock, 1992; van Merriënboer, 1990). In general, completing missing information in worked examples has been shown to increase learning and transfer in novices relative to studying complete worked examples (Otieno, Freiburg, Schwonke, & Renkl, 2011; Renkl, Atkinson, & Große, 2004; Renkl, Atkinson, Maier, & Staley, 2002; Renkl, 2002; Schwonke et al., 2009; Schwonke, Renkl, Salden, & Aleven, 2011; Schworm & Renkl, 2006). Studies vary in what information is left incomplete for learner to fill in. These completion requirements encourage the learner to process the material in a more focused and meaningful way.

One promising approach is to leave incomplete information that must be inferred from other parts of the worked example. The learner must engage in *mapping* and *integrating* information across different parts of the worked example in order to complete the missing information (Atkinson, Renkl, & Merrill, 2003; Hilbert et al., 2008; Renkl et al., 2004; Renkl, 2002; Stark, 1999). These incomplete aspects (i.e., *gaps*) require the learner to search across the other representational features of the worked example (ranging from other solution steps, the problem statement, diagrams, or databanks), interpret their meaning, and map this meaning onto the current step in the problem solving process. In other words, they must infer the goal of the operators in the current step. Thus, incomplete worked examples can facilitate links between the *goals* that the *operators* within the procedural problem-solving steps accomplish.

Despite these benefits, incomplete worked examples are not always more effective than fully worked examples. This is likely true because in these studies, the incomplete information was not focused on the central idea of integrating knowledge of goal-operator combinations (Hilbert et al., 2008; Schwonke et al., 2011). For example, gaps that focused on what stage of the proving process a geometry proof was in actually hindered learning relative to study of the same worked example without these gaps (Hilbert et al., 2008). One explanation for these results is that identifying the phase of the proving process is not central to the learning domain of proof generation.

The effectiveness of incomplete worked examples should depend on what is incomplete, although past research has not tested this directly. I propose that gaps should direct the learner to focus on aspects of the worked example that are central, not peripheral, to understanding. Specifically, gaps should foster an understanding the *goal* of the problem-solving step, and how the values and *operators* within it achieve this *goal*.

The current study contrasts incomplete worked examples with two different types of incomplete information. In one condition, the incomplete information focuses the learner on the goal-operator combinations of particular problem-solving steps; ideas that are central to learning. These gaps can be completed by integrating information across several aspects of the worked example, such as the dataset, supplemental definitions, or other worked example steps. This is accomplished through missing *intermediate* values within problem-solving steps. In the other condition, the gaps require the processing of information that can be considered peripheral to learning. The incomplete information can be readily determined by looking only within the immediate problem-solving step, and the goal need not be clear to the learner to be successfully completed. This is accomplished through missing *final* values within problem-solving steps. I hypothesize that completing *intermediate* gaps within worked examples will lead to stronger learning outcomes than completing *final* gaps that do not facilitate processing of goal-operator combinations.

The Current Study

The guiding idea of the current study is that instructional scaffolds that elicit a higher quality of focused processing on concepts *and* procedures when studying worked examples should elicit more effective learning outcomes. This study contrasts the learning benefits of instructional materials that scaffold *more* and *less* focused processing. The current study modified the level of focused processing by manipulating (a) whether students receive prompts to self-explain before receiving instructional explanations or are prompted to simply copying or paraphrase the instructional explanations and (b) whether the missing values within problem-solving steps are *intermediate* or *final* values that do and do not focus on goal-operator combinations, respectively.

Learning Context

The learning domain for this study was analysis of variance. Understanding analysis of variance is often the summative lesson within introductory statistics, and provides a strong foundation to understanding and interpreting statistics in general. It is often reported as one of the most challenging topics to teach to introductory statistics students (Gelman, 2005). For this reason, analysis of variance is an exemplary domain for investigating improvements in instructional practice.

Analysis of variance was also an ideal topic for the current study in regard to the students' knowledge level. At this point in the course the students had sufficient conceptual background knowledge of inferential statistics, and had sufficient procedural skills to calculate and understand all the subcomponents of an ANOVA calculation. However, the high-level idea of partitioning variance and using it to infer effects was a new idea, as were the specific computational procedures for doing so. The first lesson topic was one-way ANOVA, and the second lesson topic was two-way ANOVA. The learning materials were worked examples for calculating an ANOVA from raw data.

This research was conducted in a classroom context, and the manipulation was a part of normal course activities. Many of the ideas and theories this study is based on are a result of laboratory work. However, research within classrooms is important as well. Models for practical problem-solving in real-world contexts could not be developed "without reiterative cycles of both laboratory and non-laboratory based studies" (pp. 37, Scribner, 1984).

Experimental Conditions

In both conditions, students worked through a worked example of how to calculate an ANOVA. The example was broken down by sub-goal, with intermittent self-explanation prompts that pertained to the calculations that were just preformed.

More Focused Processing Condition. In the More Focused Processing Condition (*moreFP*), students were required to a) determine missing intermediate values within problem-solving steps that linked goals and operators, and b) generate self-explanations before receiving instructional explanations.

In this condition, *intermediate* values within the problem-solving steps were blank. For example, in the worked solution steps for the mean square among groups, they have to fill in MS_{ag} = _____ / _____ = *15.125*. In order to complete this, the learner has to refer to other aspects of the worked example, such as the generic algebraic formula for MS_{ag} (SS_{ag}/df_{ag}), the dataset (e.g. to find the correct value of the df_{ag}), the definition bank (e.g. what the mean square among groups represents), and the previous worked solution steps (e.g. the resulting value for SS_{ag}). In doing so, the learner determines and assigns semantically meaningful values to the gaps in the solution step that relate the operators (i.e., what values are the numerator and denominator in this particular solution step) and the goals (i.e., what these values represent, and how the operation of division relates to the meaning of the resulting MS_{ag} value) of the particular step.

After determining the intermediate values for a few sub-goals, the students respond to a set of self-explanation prompts that pertain to the problem-solving steps they just completed. When they complete working through the worked example and all the self-explanation prompts, they are given the complete worked example and instructional explanations (i.e. the 'answers' to the gaps and explanation prompts). Students are then asked to go back and *correct* their original gap values, and *correct, edit, or modify* their self-explanations so that they are correct and contain all relevant information contained within the instructional explanations.

Less Focused Processing Condition. In the Less Focused Processing Condition (*lessFP*), students are required to determine missing final values within problem-solving steps that did not link goals and operators, and receive instructional explanations and then copy or paraphrase them.

In this condition, *final* (rather than intermediate) values of the worked example steps are blank. For example, in the worked solution steps for the mean square among groups, they have to fill in $MS_{ag} = 48.4 / 3.2 =$ ____. In order to complete this, the leaner can focus on the other values within the specific step of the worked example to calculate the final value. While processing is occurring, it does not require the learner to assign meaning to the values or process the relationship between the procedural operators and the goal each problem-solving step achieves.

After determining the final values for a few sub-goals, the students read a set of instructional explanations that pertain to the problem-solving steps they just completed. When they complete working through the worked example and read all the instructional explanations, they are asked to correct their final values within the worked example and copy or paraphrase the instructional explanations to foster processing the content of the instructional explanations. This provides a rigorous control condition.

Hypotheses

I hypothesize that the More Focused Processing (moreFP) learning activities will support greater learning gains in both knowledge of procedures and concepts than the Less Focused Processing (lessFP) learning activities. These learning gains should be evident at both immediately and several days later.

CHAPTER II

METHOD

Participants

Participants were 74 undergraduate students enrolled in an Introductory Statistics course in psychology at Vanderbilt University. This was most students' first exposure to statistics, so students could reasonably be considered novices. Because this was a part of normal classroom practice, securing participant consent was not needed. Instead, IRB exempt-status approval was secured.

Students were randomly assigned to condition. There were 36 students in the moreFP condition (28 female, 8 male), and 38 students in the lessFP condition (28 female, 10 male). Students were assigned to the same condition for the entire study. There was an equal distribution of students in each class year between conditions (moreFP: 15 first year, 13 sophomore, 4 junior, 4 senior; lessFP: 14 first year, 14 sophomore, 7 junior, 3 senior). The students had similar performance prior to the occurrence of the manipulation on both exam one (moreFP: M = 153.2 (sd = 15.8) vs. lessFP: M = 145.3 (sd = 32.0), $F(1,72) = 1.732$, $p = .192$) and exam two (moreFP: M = 190.6 (sd = 48.1) vs. lessFP: M = 188.8 (sd = 50.3), $F(1,72) = 0.026$, $p = .872$).

All students participated in the study, as it was a part of regular classroom activities. Due to this, there were some absences and consequently some participants were dropped from some analyses due to missing data. There were no differences in exam one and two scores between participants who completed all activities (N=63; lessFP = 31, moreFP = 32) and those who did not (N = 11) ($F(1,72) = 0.19$, $p = .665$; $F(1,72) = 0.004$, $p = .947$).

Sample Size & Power

An a priori power analysis was conducted to determine the ideal sample size. Assuming a moderate effect size of Cohen's $D = 0.45$, an alpha level of 0.05, a power level of 0.80, and a correlation between repeated measures of 0.5, a sample of 120 participants would be required to detect a condition difference. Assuming a large effect size of Cohen's $D = 1.13$ (as found in Fonesca & Chi, 2009), a sample of 22 participants would be required. Thus, the current sample is large enough to detect a large effect but is underpowered to detect a moderate effect.

Research Design

Students participated in a pretest, intervention, immediate posttest, and delayed retention test for two different topics. During the intervention, all students worked through packets that contained worked examples and instructional explanations. Students were randomly assigned to one of two conditions: (a) Less Focused Processing (*lessFP,* $N = 38$) received incomplete worked examples that had missing final values (do not focus on goal-operator combinations) and were asked to copy or paraphrase instructional explanations, and (b) More Focused Processing (*moreFP*, $N = 36$) received incomplete worked examples that had missing intermediate values (do focus on goal-operator combinations) and self-explained before they received instructional explanations. All activities occurred during 2 weekly discussion sessions of the course.

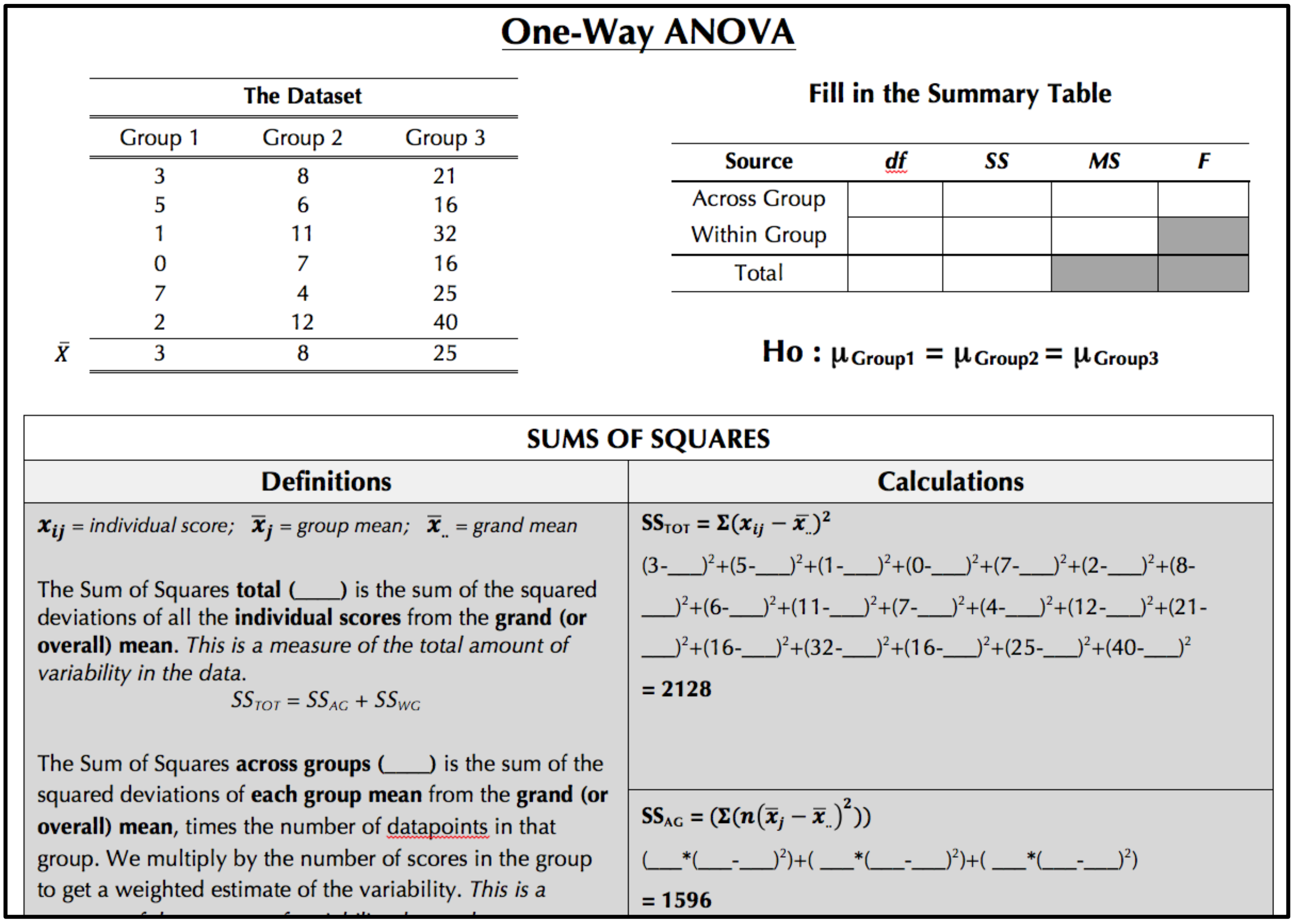
Materials

The topic for week one was one-way ANOVA, and for week two was two-way ANOVA. For each topic, intervention worksheets and assessments were developed, including a pretest, immediate posttest, auxiliary surveys, and retention test. All materials can be found in Appendix A.

Intervention Materials

Intervention worksheets were based on the existing instructor-created worksheets already planned for that week. The existing worksheets were incomplete worked examples. The worked examples guided students through calculating an ANOVA from raw data. First they calculated the sum of squares, then the mean squares, then the appropriate *F*-values, and were brought through the interpretation process. Missing values in the calculations needed to be filled in by students. In collaboration with the instructor, I wrote accompanying instructional explanations to help facilitate understanding of the worked examples, which were given to students in both conditions.

The worksheets were modified into an less focused processing version that required the learners to complete final values within the worked examples and to read and copy instructional explanations, and a more focused processing version that required the learners to complete intermediate values within the worked examples and to first generate self-explanation and then correct their explanations using the instructional explanations.



One-Way ANOVA

The Dataset			
	Group 1	Group 2	Group 3
	3	8	21
	5	6	16
	1	11	32
	0	7	16
	7	4	25
	2	12	40
$\bar{X}$	3	8	25

Fill in the Summary Table

Source	*df*	*SS*	*MS*	*F*
Across Group				
Within Group				
Total				

Ho : $\mu_{Group1} = \mu_{Group2} = \mu_{Group3}$

SUMS OF SQUARES

Definitions	Calculations
x_{ij} = *individual score*; $\bar{x}_j$ = *group mean*; $\bar{x}_{..}$ = *grand mean* The Sum of Squares **total** (____) is the sum of the squared deviations of all the **individual scores** from the **grand (or overall) mean**. *This is a measure of the total amount of variability in the data.* $SS_{TOT} = SS_{AC} + SS_{WG}$	$\mathbf{SS_{TOT}} = \Sigma(x_{ij} - \bar{x}_{..})^2$ (3-___)^2+(5-___)^2+(1-___)^2+(0-___)^2+(7-___)^2+(2-___)^2+(8-___)^2+(6-___)^2+(11-___)^2+(7-___)^2+(4-___)^2+(12-___)^2+(21-___)^2+(16-___)^2+(32-___)^2+(16-___)^2+(25-___)^2+(40-___)^2 **= 2128**
The Sum of Squares **across groups** (____) is the sum of the squared deviations of **each group mean** from the **grand (or overall) mean**, times the number of datapoints in that group. We multiply by the number of scores in the group to get a weighted estimate of the variability. *This is a*	$\mathbf{SS_{AC}} = (\Sigma(n(\bar{x}_j - \bar{x}_{..})^2))$ (___*(___-___)^2)+(___*(___-___)^2)+(___*(___-___)^2) **= 1596**

Figure 1. Worked Example Layout from MoreFP Condition

Worked Examples. Many features of the worked examples were the same across conditions. The worked example consisted of several parts. There was a problem statement, a data set, worked example steps presented by sub-goal, the actual worked example steps with values from the problem statement, and a summary table to collect the final values (see Figure 1). The definition sections had blanks that all learners were asked to fill in. These were very simple blanks to ensure the learner attended to and could reproduce the information contained within the definition (e.g. The abbreviation for the mean square within groups is: __[MSwg]__). The worked examples had subgoals grouped together, such as sum of squares total, sum of squares among group, and sum of squares within group. Subgoal grouping has been shown to be effective in supporting novice learning (Atkinson et al., 2000; Catrambone, 1998).

All aspects of the worked example were the same across conditions, with the exception of which values were missing within the problem-solving steps in the calculations section.

LessFP Worked Examples. The gaps within the worked example required the learner to calculate the resulting final value for particular steps within the worked example. All the information necessary to correctly complete the gap was contained locally within that particular problem-solving step. In part two these correct values were filled in.

MoreFP Worked Examples. These worked examples also contained gaps, but these gaps were of intermediate values within the solution step. In order to correctly fill in these gaps, the learner would have to gather and integrate information from various sources within the worked example in the service of combining goals and operators. In part two these correct values were filled in.

<table>
<tr><th colspan="2">STATISTICAL CONCLUSION</th></tr>
<tr><th>Definitions</th><th>Calculations</th></tr>
<tr><td>This is the critical value by which we judge if the amount of variability between groups is less than 5% likely to occur between groups from the same population.</td><td>$_{.05}F_{\_\_,\_\_} = 3.68$ (from table)
Decision: <u>Reject</u> Fail to Reject</td></tr>
<tr><th colspan="2">Thought Question</th></tr>
<tr><td colspan="2">• Interpret your conclusion in terms of the treatment effect. Discuss variability among groups and within groups.
There is a treatment effect, and not all of the group means are the same. The variance among groups was much greater than the variance within groups. This means that the variability due to differences of group means (the treatment effect and noise) was greater than the variability of individual scores from their own group mean (noise). So we can conclude that there is a treatment effect.</td></tr>
</table>

Figure 2. An Instructional Explanation

Instructional Explanations. After each grouping of subgoals of the worked example, there were supplemental instructional explanations. The instructional explanations were framed as 'thought questions' (explanation prompts) and 'answers' (either instructional- or self-explanations). See Figure 2. They provided information about the goal of the particular operations (Gerjets et al., 2006) and concepts, which are rationales underlying specific problems (van Gog et al., 2006; Wittwer & Renkl, 2010) (e.g. "*How does the F-ratio quantify the treatment effect? The F-ratio divides the* MS_{AG} *by the* MS_{WG}. *The* MS_{WG} *contains the variability due to the treatment effect and noise...*" and "*What main idea are these Sums of Squares capturing? The main idea is that we are capturing the amount of variability in the data...*"). Both types of instructional explanations have been used effectively in prior research, with meta-analytic effect sizes of 0.92 and 0.17, respectively (Wittwer & Renkl, 2010). In both conditions, participants studied the same instructional explanations. However, the way in which the instructional explanations were presented was manipulated between conditions.

Intervention Worksheet Answer Keys. At the end of each intervention session, both parts one and two were collected for analyses. Students were given an answer key for the worksheet they completed during the intervention. These were for them to study from, as was in line with typical class practice.

Assessments

The assessments were designed to tap the students' understanding of both procedural and conceptual aspects of analysis of variance. An understanding of procedures entails knowing the formulas and values needed to successfully execute action sequences for problem solving (Anderson, 1993; Rittle-Johnson & Alibali, 1999). An understanding of concepts entails an understanding of principles governing a domain and the interrelations between units of knowledge (Bisanz & LeFevre, 1992; Greeno, Riley, & Gelman, 1984; Rittle-Johnson, Siegler, & Alibali, 2001). These materials were designed in collaboration with the course instructor and content-area experts. The instructor, two fifth-year Quantitative Psychology PhD students, and three undergraduate research assistants who have already completed the course provided input to develop and refine these materials.

Pretest. The pretest items captured the amount of pre-existing knowledge the student had of the topic, and was designed to take fewer than 5 minutes to complete. The items asked: 1) At a high level, how does an one–way ANOVA determine if there are differences between groups? 2) Briefly list the steps needed calculate an one–way ANOVA; and 3) How do the steps involved in an one–way ANOVA answer the question we are trying to test? The pretest items for week two were the same but one-way ANOVA was replaced with two-way ANOVA.

Posttest Items. The immediate posttest contained items that were adapted from the course textbook, other textbook series and assessment resource books. They tapped an understanding of the concepts and procedures that underlie ANOVA that were outlined in the intervention activity. Items that focused on knowledge of procedures required various modifications of the procedure (e.g. determining the SS_{WG} by subtracting SS_{AG} from SS_{TOT} instead of calculating it from the raw data). All posttest procedural items therefore could be considered procedural transfer items. The conceptual knowledge items tapped participants' ability to explain the principles that underlie the example they just studied (Wittwer & Renkl, 2010), (e.g. *When the amount of variability within groups increases, but the amount of variability among groups stays the same, the value of the F-ratio (increases/decreases). Why?*). Items required that the learner make a conceptual inference, provide justifications, or apply or adapt a problem solving procedure. Response formats included short answer, multiple choice, or numeric answers. See Appendix A for the complete assessment. The week one posttest contained 15 items and the week two posttest contained 28 items.

Retention Test Items. The retention items were similar to those on the posttest, but not identical. The week one retention contained 20 items, and the week two retention contained 25 items.

Auxiliary Survey Measures

The learners were presented with Likert-scale rating items that assessed self-reported cognitive load, depth of processing, and perceived helpfulness. These measures were designed as manipulation checks.

Cognitive Load Scale. Cognitive load refers to any demands on working memory storage and processing of information (Schnotz & Kürschner, 2007). These items tapped the amount of cognitive load the learner experiences during learning. This was used as a metric of learners' cognitive processing activity. Seven items were adapted from the subjective rating of mental effort and of task difficulty by (Paas, 1992), such as "In solving or studying the preceding problem I invested: (very low to very high) mental effort." A total cognitive load score was calculated by determining the average rating for each participant, and then comparing these scores between conditions. The moreFP condition was expected to report higher cognitive load.

Active and Constructive Activity Scale. These self-report items were used to identify level of 'active' and 'constructive' activity undertaken by the learner, following the definitions set forth by Chi (2009). An *active* activity requires the learner to select or manipulate the learning materials. A *constructive* activity requires the learner generate new information beyond what is provided in the learning materials. It is hypothesized that constructive learning activities result in superior learning gains than active activities. These items were adapted with permission from Bujak (2010). This framework posits that activities included in the moreFP condition require more constructive activities than those in the lessFP condition (who utilize primarily active activities) (Chi, 2009). As such, these items were used as a measure of amount and type of cognitive processing undertaken by the participants. Statements about active or constructive cognitive processing were presented, such as “I identified the most important ideas” (active) or “I connected the text to ideas I already know” (constructive). Participants were asked to rate the frequency with which they engaged in each of these activities either overtly or covertly on a 5-point Likert scale. There were 6 constructive items and 5 active items, and separate scores on the two scales (constructive and active) were calculated by determining the average rating on each scale for each participant. These scores were then compared between conditions. The conditions were expected to have equal levels of ‘active’ processing, and the moreFP condition was expected to have higher levels of ‘constructive’ processing.

Perceived Usefulness Scale. These 6 items tapped the amount of perceived usefulness of the worked examples and instructional explanations. These were included because often learners perceive instructional explanations to be very helpful, even if they are not as useful for increasing knowledge as self-explanation prompts (Schworm & Renkl, 2006). These items concern the perceived usefulness of the learning environment and the subjective learning outcomes. The items were answered on a Likert scale from 1 (does not apply at all) to 6 (totally applies), as in Schworm & Renkl (2006). The items included statements such as 'These calculations helped me to understand ANOVA'. Other items tapped subjective usefulness of the materials such as 'I will think of these thought-questions when solving ANOVA problems in the future'. These items achieved good reliability in prior work. Students in the two conditions provided similar ratings in week one. Due to time constraints, these items were omitted from week two materials. As the conditions did not differ in their week one ratings, this scale will not be discussed further.

Procedure

This study took place during the last third of the course and was primarily carried out during 2 lab sections. Students attended professor-led lectures twice a week, and met in smaller teaching assistant-led lab sections of about 12 students once a week. The three teaching assistants were advanced graduate students in psychology, including myself. During the lab sections, the students worked through worksheets and assessment problems on a different topic each week. On each 50 minute lab section day of the study, students completed a pretest, the intervention activity, the immediate posttest and auxiliary surveys. A retention test was administered in class 4 days later. The first and second lab sections occurred 3 weeks apart due to a university-wide holiday break.

Instructions were written on the board, along with timeline guides for each section. A large digital clock was displayed on a computer monitor for the students, and all students were asked to note the time when they began and finished each part of the study materials. In line with the activities that took place during the normal course lab sections, work was not graded but students were told to do their best. All students brought and used calculators. See Appendix B for intervention day teaching assistant scripts.

Students completed the pretest, turned it in, and then began on the part one worksheet. In part one, students were asked to complete the missing values within the worked example and read instructional explanations.

LessFP. In part one, lessFP participants were asked to complete the fill-in-the-blanks in the definitions and calculate the final value of each step of the worked example. They also received instructional explanations in part one and were asked to read them carefully. When a student completed part One, they were given part two and a purple pen. They students were asked to use the purple pen for all their work from then on, because correcting or modifying their work on part one would be required and this distinction was important for analyses. In part two, the instructional explanations were blank, and the students were asked to copy or paraphrase the instructional explanations from part one into part two. This ensured the student attended to and processed the content of the instructional explanations.

MoreFP. In part one, only the thought questions were presented, and students were prompted to generate self-explanations as ‘answers’ to the thought questions. When a student completed part One, they were given part two and a purple pen. They students were asked to use the purple pen for all their work from then on, because correcting or modifying their work on part one would be required and this distinction was important for analyses. Part two contained the instructional explanations, which were answers to the 'thought questions' (see Figure 2). The students were asked to read the instructional explanations and then go back and *edit, modify or improve* their original explanation responses with the provided purple pen so that it contained all the main ideas presented within the instructional explanations.

Once the student completed and turned in both parts one and two, the students then completed the immediate posttest and auxiliary surveys. The next day of class, four days after the intervention activity, the students were given a retention test. The students were told this was to see how much they learned and remembered from the intervention day, and that their responses would not be graded. Administration took about 12 minutes.

Missing Data

Due to the real-world nature of this study and typical absences, not all participants were present for all aspects of the study. Only participants who completed the intervention and the retention test were included in the analyses. The pattern of findings was the same when analyses on the immediate posttest included all participants. See Table 1 for participant completion information.

Component	N	LessFP	MoreFP
Week One			
Intervention	72	36	36
Retention	62	31	31
Week Two			
Intervention	64	31	34
Retention	62	30	32

Table 1. Participant Completion Counts

Coding

Intervention Materials. Participant work on the intervention items was coded for worked example accuracy and explanation accuracy and quality.

Worked example accuracy was coded. All blank aspects of each worked example were coded for a *correct numeric* response. The percentages of correctly filled-in missing values within the worked examples were calculated. Part one worked example accuracy was the percentage of initially correct filled-in values. This was determined by dividing the original number of correctly completed values by the number of total missing values. Part two worked example accuracy was the percentage of corrections the participants made after receiving the answer key in part two. This value was determined by dividing the number of values the participant corrected while checking their work by the number of total missing values. Final worked example accuracy was the overall percent of correct values after correction. This was determined by dividing the number of correctly filled in values across both parts of the intervention activity by the number of total missing values. For example, if a participant had a part one worked example accuracy score of 50%, a part two worked example correct score of 25%, their final worked example accuracy score would be 75%. Recall that the lessFP participants completed missing final values, whereas the moreFP participants completed missing intermediate values.

The intervention explanation responses were evaluated. Each participant's percent overall explanation quality score was calculated. All codes were assigned according to a criterion-based rubric. Percent valid explanation responses were determined by coding the participants' responses as either invalid (0), valid (1), or high quality (2). To be considered high quality, the response had to include the one or two central idea contained within the correct explanation response, and at least 75% of all other supporting points. The percentage overall explanation quality score was the sum of the explanation scores divided by the maximum possible score. For the lessFP condition, the copied or paraphrased explanations at part two were coded. For the moreFP condition, explanations were coded at part one (self-explanations) and part two (corrected self-explanations).

Completion & Time. Metrics of amount of time spent and subjective level of completion were collected for each aspect of the intervention activity. On each section of the intervention, students were asked to list their start and stop times. Students were also asked if they finished the activity, and if they would have wanted more time. Completion times in minutes and seconds and self-report completion rates are reported in Figure 7.

Assessment Items. Assessment items were coded as either valid or invalid according to a criterion-based coding scheme. Items that had a single numeric answer were coded as valid if the number was correct. The 8 short answer items that required more subjective coding were coded using a 2 point scale– 0 if invalid, 1 point if valid but low quality, and 2 points if valid and high quality. This 2-point coding scheme was used to increase measurement sensitivity to the range of response qualities. Scores on each assessment were calculated by summing all the earned points on an assessment and dividing it by the possible number of points, resulting in a percent correct score. Total number of possible points are as follows: week one posttest and retention had 15 and 20 points, respectively, and week two had 25 and 19, respectively.

Inter-rater Reliability. A reliability analysis was performed on the short-response assessment items because they required subjective coding. A subsample of 20% of the participants had their responses to these items double-coded by another researcher. Kappa coefficients ranged from .617 (substantial) to .944 (almost perfect; (Landis & Koch, 1977)), with one exception. One conceptual item on the week two posttest had only slight agreement due to the varied and often vague nature of student responses. This item was dropped from further analyses.

Instrument Screening

Item Analyses. Item-level analyses were preformed to evaluate the informativeness and fit of each item. Analyses were completed using all participants using both classical test theory and item response theory measures. Items were analyzed within the context of its assessment; for example, week one pretest items were evaluated relative to all week one pretest items as a whole. The items that utilized a two-point coding scheme had two entries for this analysis (one point if it was valid, and another point if it was high quality), as binary data was required. See Appendix C for item metrics and flagged items.

The item-level analysis was used to flag items that may be uninformative or problematic. The following criteria were used to evaluate poor item fit: (a) if the item-total correlation was low (< 0.1) or negative, (b) if the item had poor outfit metrics (>2 or <.5, (c) if the item had poor infit metrics (>1.5) or (d) if it had a low mean score (<.10) . Items that had been flagged by multiple indices were considered for removal from the scales. Final exclusion decisions were based on an evaluation of individual item's fit scores, with more weight given to the more informative indices (a) and (b). Dropped items in week one were 1 conceptual posttest item and 6 procedural retention test items. Items dropped from week two were 3 conceptual posttest items and all 4 conceptual retention test items. All outcome analyses were conducted with these items dropped.

Final assessment items for week one were 3 pretest items, 15 posttest items (7 conceptual and 8 procedural), and 20 retention test items (5 conceptual and 15 procedural). For Week 2, there were 3 pretest items, 25 posttest items (11 conceptual and 14 procedural), and 19 retention test items (0 conceptual and 19 procedural).

Instrument Reliability. Instrument reliability, after items with poor fit were dropped, was evaluated using Cronbach's alpha. Analyses were done by assessment time point, and also by subscale. The items that utilized a two-point coding scheme had two entries for this analysis, as binary data was required. A Cronbach's alpha value of 0.7 or above is acceptable. Only the week one pretest had an alpha value below 0.7, and most were above 0.8 (see Table 2).

Assessment Component	N	Scale Alpha
Week One		
Pretest	6	0.624
Posttest	20	0.862
Conceptual	12	0.766
Procedural	8	0.887
Retention	22	0.931
Conceptual	7	0.820
Procedural	15	0.913
Week Two		
Pretest	6	0.767
Posttest	27	0.934
Conceptual	13	0.844
Procedural	14	0.934
Retention	21	0.950
Conceptual	0	na
Procedural	21	0.950

Table 2. Assessment Reliability.

CHAPTER III

RESULTS

First, information about the pre-existing measures about the participants by condition is presented. Second, information about intervention performance is presented, including intervention compliance, accuracy, and cognitive engagement levels. Finally, analysis of the outcome assessments performance is presented. All analyses include only the participants who completed the respective week's intervention activity and retention test.

Pretest Knowledge

As shown in Table 7, there were no differences between conditions in pre-existing knowledge of the week one intervention topic, according to week one pretest scores, ($F(1, 72) = 1.696$, $p = .197$). However, there were differences between conditions on the week two pretest, with the moreFP condition scoring significantly higher ($F(1, 63) = 4.425$, $p = .039$, $\eta^2_p = .066$). This suggests that differential learning between conditions may have occurred during week one that influenced performance on week two.

Additionally, scores on the week two pretest were significantly higher than week one ($F(1,122) = 4.681$, $p = .032$, $\eta^2_p = .037$), suggesting that learning the week one material benefitted performance at week two. This makes sense, considering that the topics of One- and Two-Way ANOVA build on highly similar concepts and procedures.

Intervention Activities

Level of intervention compliance was evaluated through the participants' accuracy within the worked examples, their explanation response quality, their completion self-report, and their level of cognitive activity self-report. The following analyses were done using an ANOVA model with dependent variable being the intervention activity in question, and with condition as the independent variable. Analyses were conducted separately by week, with separate results for week one and then for week two. This was then followed by a between-week analysis with condition collapsed, the intervention activity in question as the dependent variable, and week as the dependent variable. The between-condition analyses include only the participants who completed the respective week's intervention activity, and the between-week analyses include only the participants who completed both weeks' intervention activities.

Worked Example Accuracy

Worked example accuracy was examined between conditions. Part one worked example accuracy did not differ between conditions at week one or at week two. Part two worked example correction scores were significantly higher in the moreFP condition in week one ($F(1,122) = 4.039$, $p = .047$, $\eta^2_p = .032$), but not week two. Because of this, final worked example accuracy was higher in the moreFP condition than the lessFP condition in week one (93% vs. 82%, $F(1,70) = 5.103$, $p = .027$, $\eta^2_p = .068$), but not week two. See Table 3.

The level of worked example accuracy varied between weeks when condition was collapsed. Part one worked example accuracy was significantly higher at week one than week two (79% vs. 67%; $F(1,122) = 8.310$, $p = .005$, $\eta^2_p = .064$). Part two worked example accuracy did not differ between weeks. Final worked example accuracy was higher at week one than week two (89% to 78%, $F(1,122) = 13.452$, $p = .000$, $\eta^2_p = .099$).

To summarize, between conditions, the moreFP condition was more successful in completing the gaps in the worked example than the lessFP condition at week one, but not at week two. Between weeks, the part one and final worked example accuracy scores were higher in week one than in week two, suggesting that the participants may have invested less effort in completing the worked examples in week two.

Explanation Quality

The participants' explanation responses were evaluated as an additional metric of intervention compliance. All analyses were performed using participants' part two explanation quality score, which was the lessFP participants' copied or paraphrased instructional explanations and the moreFP participants' corrected self explanations. See Table 3.

Intervention Compliance Scores					
		LessFP	*sd*	MoreFP	*sd*
Week One					
Part One	Expln Quality			45%	*20%*
	Worked Ex Acc	76%	*24%*	80%	*27%*
Part Two	Expln Quality	75%	*26%*	76%	*20%*
	Worked Ex Acc	6%	*12%*	13%	*20%*
Week Two					
Part One	Expln Quality			31%	*21%*
	Worked Ex Acc	70%	*18%*	66%	*25%*
Part Two	Expln Quality	67%	*25%*	63%	*27%*
	Worked Ex Acc	7%	*11%*	13%	*18%*

Table 3. Intervention Compliance Scores

Participants' explanation quality was examined between conditions. When considering part two explanation quality, both conditions had equal scores at both weeks one and two (see Table 3).

Explanation quality was compared between weeks, with condition collapsed. Both conditions' part two explanation quality scores were higher in week one than week two (76% to 64%; $F(1,122) = 7.68$, $p = .006$, $\eta^2_p = .059$). When considering part one explanation quality, which only concerns the moreFP condition, their scores were much higher in week one when compared to week two (47%, sd = 19% to 30%, sd = 20.2; $F(1,64) = 11.822$, $p = .001$, η^2_p =

.156). Generally, explanation quality was higher at week one than week two.

Additionally, the moreFP participant's explanation quality was also examined between part one and part two. There were differences between the moreFP's initial generated self-explanations at part one and their corrected versions at part two, indicating that learning from the instructional explanations occurred. Descriptively, explanation quality improved from part one to part two at week one (45% to 76%), and week two (31% to 63%). Being presented with instructional explanations and having the opportunity to correct original self-explanation responses improved explanation quality.

In sum, there were no differences in part two explanation quality between conditions. When contrasting between week one and week two, explanation quality was higher at week one, as were the moreFP participants' part one explanations. Additionally, the moreFP students explanations were significantly improved from part one to part two in both weeks, suggesting that learning from the instructional explanations occurred. Taken together, this suggests that participants expended more effort in their explanations in week one.

Intervention times and self-reported completion

Completion time for each aspect of the intervention was collected. Both conditions took the same overall amount of time for the pretest, the intervention, and the posttest quiz on both weeks. However, the conditions differed in completion time between intervention part one and part two. Completion times were analyzed between conditions with an ANOVA model separately for each week. At week one, the moreFP condition used more time to complete part one ($F(1,70) = 35.286$, $p = .000$, $\eta^2_p = .335$); and the lessFP condition used more time to complete part two ($F(1,62) = 52.289$, $p = .000$, $\eta^2_p = .458$). The same pattern held at week two, with the moreFP condition using more time for part one ($F(1,55) = 7.414$, $p = .009$, $\eta^2_p = .119$), and the lessFP condition using more time for part two ($F(1,48) = 11.930$, $p = .001$, $\eta^2_p = .199$).

This difference, however, was by design, and the total amounts of time to complete both parts of the intervention were equal between conditions. See Table 4.

Completion Times by Condition in Minutes			
Component	LessFP	MoreFP	Average
Week 1			
Pretest	03:59	03:40	03:49
Part One	11:16	15:52	13:34
Part Two	10:53	07:04	09:02
Posttest	13:25	12:51	13:08
Week 2			
Pretest	03:14	03:11	03:12
Part One	20:14	23:02	21:49
Part Two	08:50	05:31	07:01
Posttest	13:09	12:50	12:59

Table 4. Study Completion Times

Participants were asked to self-report their completion level of the intervention. At the end of intervention part two, participants were asked 1) "Did you finish? Yes/No" and 2) "Would you have wanted more time? Yes/No". See Figure 3. Notice that a sizable portion of all participants reported that they did not finish the activities to their satisfaction and would have wanted more time. In week one, significantly more students in the lessFP condition reported being incomplete ($F(1,61) = 4.969$, $p = .029$, $\eta^2_p = .075$), and wanting more time ($F(1,70) = 10.449$, $p = .002$, $\eta^2_p = .130$). There were no significant differences in week two.

Completion self-reports were compared across weeks. There were no significant differences in self-report of finishing the intervention, nor in wanting more time to complete it between weeks one and two when condition was collapsed.

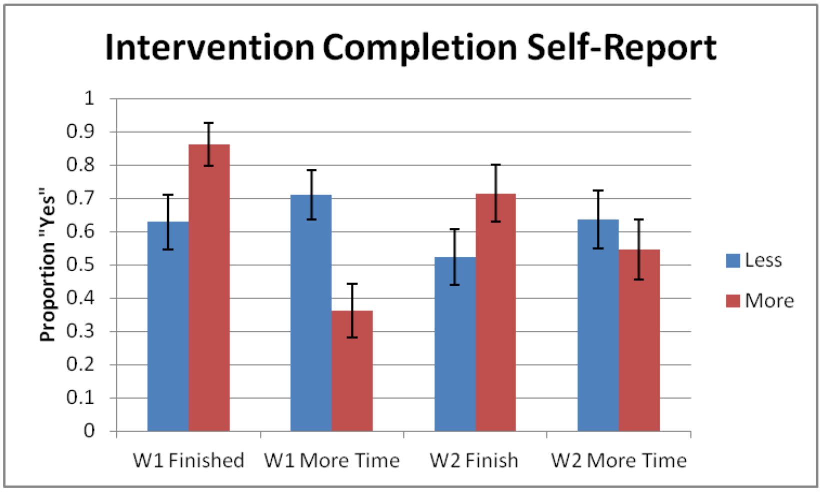


Figure 3. Intervention Completion Self-Report

Manipulation Check – Auxiliary Survey Items

Participants completed surveys after they completed the intervention worksheets and before the immediate posttest. These items tapped participants' subjective cognitive load and levels of active and constructive activity. Between condition differences were evaluated by comparing the average scores for each survey by condition using the same ANOVA models described previously. See Table 5.

Auxiliary Survey Likert Scores		
	LessFP	MoreFP
Week One		
Cognitive Load	3.47	3.74
Active	3.31	3.22
Constructive	2.15	2.96
Week Two		
Cognitive Load	3.52	4.00
Active	2.97	3.11
Constructive	2.50	3.30

Table 5. Auxiliary Survey Rating Scores by Condition

Cognitive Load Scales. In both week one and week two, the moreFP condition reported higher levels of cognitive load than the lessFP condition (week one, $F(1,61) = 3.99$, $p = .050$, η^2_p =.061; week two, $F(1,45) = 6.716$, $p = .013$, $\eta^2_p = .130$). There were no differences in amount of reported cognitive load between week one and week two when condition was collapsed.

Levels of Active and Constructive Activity. These self-report items were used to identify level of 'active' and 'constructive' activity undertaken by the learner. Both conditions reported engaging in the same amount of ‘active’ cognitive processes at both weeks. As hypothesized, the students in the moreFP condition reported engaging in more ‘constructive’ cognitive processes at both weeks (week one: $F(1,72) = 7.26$, $p = .009$, $\eta^2_p = .092$; week two: $F(1,43) = 9.926$, $p = .003$, $\eta^2_p = .188$). Between weeks, participants across both conditions reported engaging in ‘active’ learning activities more in week two than in week one ($F(1,93) = 4.609$, $p = .034$, $\eta^2_p = .010$). There were no differences in reported ‘constructive’ activities between weeks.

Assessment Outcomes

Student performance on both the immediate posttest and delayed retention test was investigated. Results are presented by week. All raw scores and standard deviations for only those who completed their respective week’s intervention activity are presented in Table 6.

Assessment Proportion Correct by Condition						
	LessFP			MoreFP		
	N	Mean	SD	N	Mean	SD
Week 1						
Pretest	37	.414	*.274*	35	.498	*.267*
Conceptual	37	.297	*.323*	35	.368	*.308*
Procedural	37	.649	*.388*	35	.757	*.335*
Posttest	37	.555	*.253*	35	.671	*.201*
Conceptual	37	.527	*.259*	35	.612	*.189*
Procedural	37	.529	*.326*	35	.670	*.259*
Retention	30	.472	*.279*	30	.592	*.249*
Conceptual	30	.479	*.283*	30	.575	*.296*
Procedural	30	.433	*.291*	30	.560	*.327*
Week 2						
Pretest	31	.519	*.325*	34	.662	*.217*
Conceptual	31	.440	*.373*	34	.585	*.246*
Procedural	31	.677	*.355*	34	.816	*.284*
Posttest	31	.574	*.233*	34	.620	*.247*
Conceptual	31	.695	*.185*	34	.690	*.220*
Procedural	31	.403	*.329*	34	.500	*.351*
Retention	27	.394	*.317*	30	.440	*.331*
Conceptual	27	.111	*.212*	30	.150	*.233*
Procedural	27	.404	*.315*	30	.447	*.331*

Table 6. Assessment Proportion Correct by Condition

Analyses and Covariates. The effect of condition on assessment outcomes was analyzed with a repeated measured ANCOVA model. Each week's post and retention test scores were used as the dependent variables. Supplementary analyses were preformed to test effects of condition on knowledge of concepts and of procedures separately.

Pretest for a given week was included as a covariate to control for differences in prior knowledge. Metrics of intervention compliance were also tested for their predictive value. Overall worked example accuracy and final explanation quality were included in exploratory models. Overall worked example accuracy predicted some outcomes, but not others. Final explanation quality always predicted outcomes. Thus, overall explanation quality was included as a covariate because intervention compliance was an important factor in how effective the intervention was.

Week One Outcomes

At week one, the moreFP condition had significantly higher outcome scores ($F(1,56) = 6.685$, $p = .012$, $\eta^2_p = .107$). See Figure 4. There was a significant effect of test time, with scores at posttest being higher than those at retention ($F(1,56) = 4.953$, $p = .030$, $\eta^2_p = .081$). There was no significant condition by test time interaction. In week one, pretest scores did not impact outcomes ($F < 1$, $p = .388$), but explanation quality did ($F = 12.495$, $p = .001$, $\eta^2_p = .182$). Follow up analyses indicated that the benefit of the moreFP condition was significant for knowledge of procedures ($F(1,56) = 7.664$, $p = .008$, $\eta^2_p = .120$) (Figure 5), and was marginal for knowledge of concepts ($F(1,56) = 3.176$, $p = .080$, $\eta^2_p = .054$) (Figure 6).

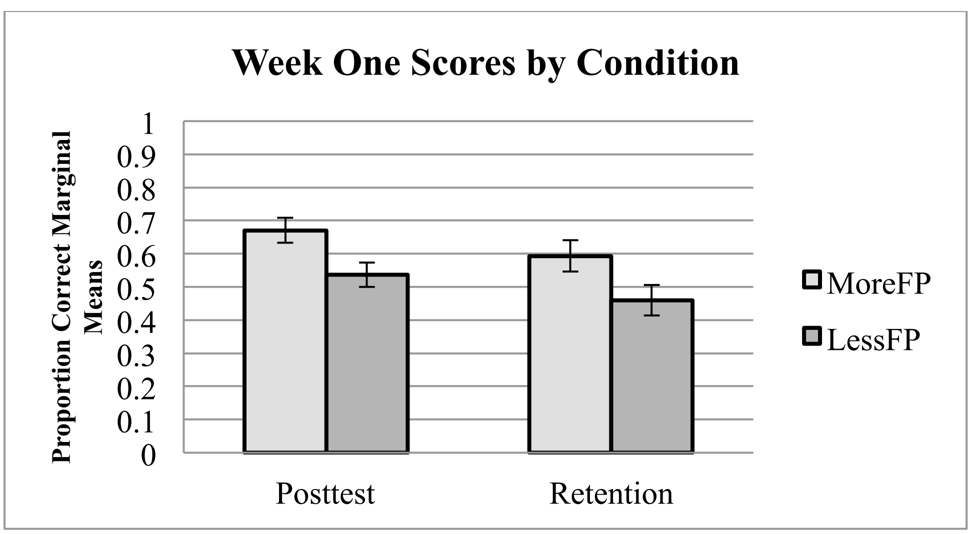


Figure 4. Week One Outcome Scores by Condition

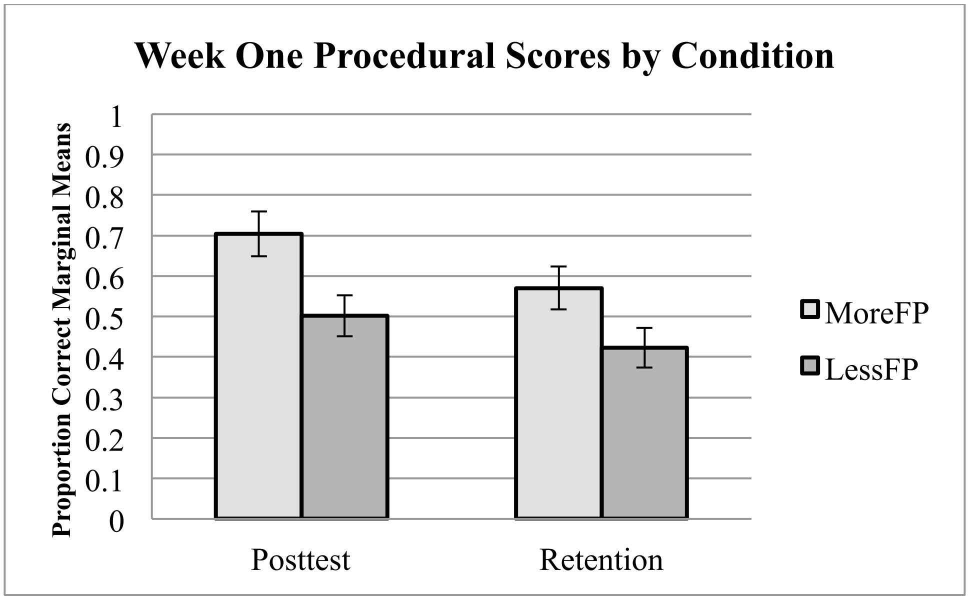


Figure 5. Week One Procedural Knowledge Outcome Scores by Condition

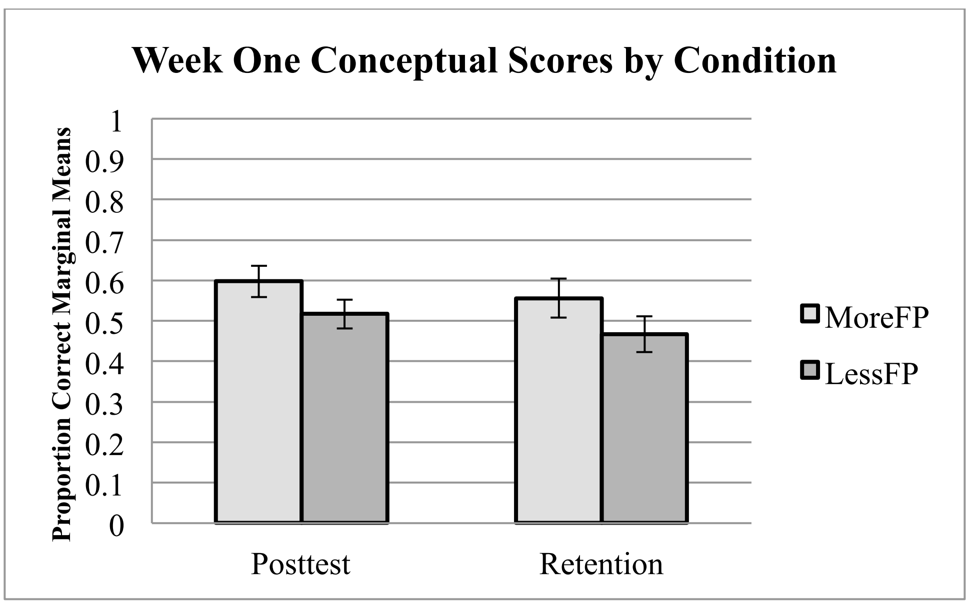


Figure 6. Week One Conceptual Knowledge Outcome Scores by Condition

Week Two Outcomes

At week two, there was no effect of condition ($F(1,53) < .01$, $p = .989$, $\eta^2_p < .001$). There was a significant effect of test time, with scores at the posttest being higher than those at retention ($F(1,53) = 7.076$, $p = .010$, $\eta^2_p = .118$). There was no significant condition by test time interaction. See Figure 7. Week two pretest was predictive of outcomes ($F(1,53) = 8.102$, $p = .006$, $\eta^2_p = .133$), whereas the final explanation quality was not ($F(1,53) = 2.4$, $p = .127$, $\eta^2_p = .043$). Follow-up analyses indicated that the conditions performed equally well for knowledge of concepts ($F(1,53) = 0.008$, $p = .927$, $\eta^2_p < .001$); and of procedures ($F(1,53) = 0.02$, $p = .888$, $\eta^2_p < .001$). See Table 6.

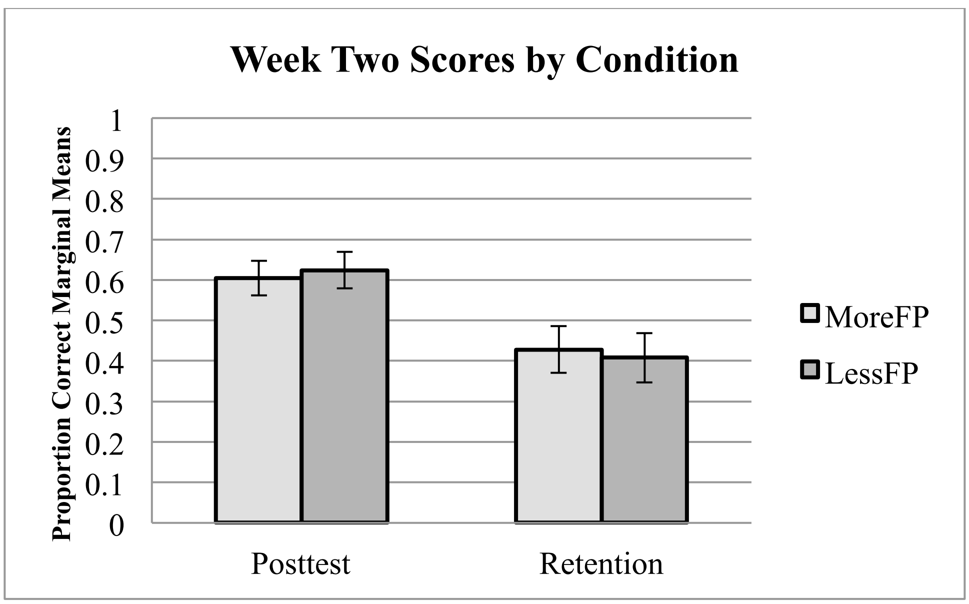


Figure 7. Week Two Outcome Scores by Condition

Summary of Results

There were no pre-existing differences in pretest knowledge at week one. However, the moreFP condition had higher pretest scores at week two. In terms of intervention compliance, worked example accuracy was higher for the moreFP participants at week one, although there were no conditions differences at week two. The worked example accuracy scores across both conditions were higher at week one than at week two. There were no differences in explanation quality between conditions. When contrasting between week one and week two, explanation quality was higher at week one. Self-reported intervention completion rates were lower for the lessFP participants at week one only. Self-reported cognitive load was higher for the moreFP condition at both weeks one and two. The moreFP condition also self-reported higher levels of constructive activity at both weeks. At posttest and retention tests, there was a benefit of the moreFP condition in week one. However at week two, there were no differences between conditions.

CHAPTER IV

DISCUSSION

This study suggests that the modified focused processing stance can benefit learning when students engage with materials appropriately. When learners are encouraged to process both concepts and procedural goal-operator combinations in a productive way, learning outcomes were often greater than engaging with the same materials in a less focused way in week one. In particular, I demonstrated that a combination of explanation prompts and incomplete worked examples that focus on goal-operator combinations with an explicit correction process can be an effective method to increase focused processing. However, student engagement in the desired activities may be difficult to sustain, as the effect of condition was no longer present in week two.

My modified focused processing stance elaborates on the focused processing stance put forth by Renkl & Atkinson (2007). Their stance states that the interactive features of a learning environment should not only elicit active processing of learning materials, but also focus attention on the primary *concepts* of the domain. However, there can be a cost to focusing attention primarily on concepts. In particular, Berthold et al. (2011) and Berthold and Renkl (2009) found that when explanation prompts focused on concepts, knowledge of concepts improved relative to not receiving explanation prompts. However, there was a simultaneous *detrimental* effect of the learning of procedures. They suggest this detriment may have been the result of the learning environment requiring too much cognitive load. Therefore the learners could not adequately process all of the learning material, so had to neglect the procedural aspects of it to fully engage with the conceptually-oriented prompts.

The modified focused processing stance overcomes these limitations by scaffolding the learner to focus on concepts *and* procedures. This suggests that too much cognitive load is not the reason why the prior conceptualization of focused processing failed. By their reasoning, the current manipulation should have elicited even greater amounts of cognitive load, thus leading to even lower learning outcomes. However, the current study found learning benefits of both types of knowledge in week one.

My modified focused processing stance overcomes the former limitations on knowledge of procedures. This modified focused processing stance is important, as an understanding of both concepts *and* procedures is necessary for a full understanding of many domains. Additionally, I more clearly specify what the learner should focus on to maximize learning outcomes.

Following is a discussion of potential processes that may have contributed to the superior learning outcomes of the more focused processing condition at week one. These potential processes include explanation generation and correction, as well as those behind linking problem solving step operations with the goals they accomplish. The relationship of the current findings to other learning frameworks is then presented. This is followed by a brief discussion of the possible reasons for no differential effect at week two, and a discussion of limitations and future directions.

Focus on Explanation Generation & Correction

A learner can increase their focused processing by first generating a self-explanation response. This generation process may activate several potential cognitive processes that benefit learning, such as activating prior knowledge, assimilating and integrating new information, increasing memory trace strength via the generation and recall process, and making new inferences (Fonseca & Chi, 2011). For example, self-explanation generation has been found to increase the memory trace of correct problem solving procedures, as well as decrease those of incorrect procedures (Siegler & Lin, 2010). The generation process may also better prepare learners for future learning from the instructional explanations that follow (Schwartz & Bransford, 1998; Schwartz & Martin, 2004). In the current study, after learners self-explained, they received a correct explanation and engaged in an explicit correction process. This ensured that the learner not only received the correct information, but also processed and integrated this correct information into their own understanding. Indeed, learners self-corrected nearly 40% of their explanations. This correction process seems to contribute to greater learning and understanding relative to viewing the same correct instructional explanation first, and then processing it via copying or paraphrasing. Indeed, previous research indicates that generating ideas prior to instructional explanations leads to greater learning than reproducing ideas after the same instruction (DeCaro & Rittle-Johnson, 2012; Schwartz, Chase, Oppezzo, & Chin, 2011). The generation and correction process in the current study likely contributed to enhanced learning.

This specification of generating and then correcting is important. Multiple studies have tried to find the optimal combination of self- and instructional-explanations, and there has not been a consensus of what this might be (Wittwer & Renkl, 2010). Several studies have failed to find a benefit for providing instructional explanations in combination with prompts to self-explain. For example, a study on effective design of learning materials with student teachers found that it was not beneficial to include instructional explanations before prompts to self-explain compared to self-explanation prompts without any prior instructional explanations (Hilbert, Schworm, & Renkl, 2004). Contrast this with a similar study on probability with undergraduates that provided instructional explanations first and found no benefit of follow-up self-explanation prompts compared to no follow-up self explanation prompts (Gerjets et al., 2006). Yet another study, also on effective design of learning materials with student teachers, found that self-explanation prompts alone resulted in the highest learning outcomes when compared to a combination of self- and instructional-explanations, which were in turn better than just instructional explanations (Schworm & Renkl, 2006). Other studies document the benefits of learner-dependent approaches, such as providing instructional explanations adaptively, only when the learner requests them (Merrill, Reiser, Merrill, & Landes, 1995; Renkl, 2002; Sánchez, García-Rodicio, & Acuña, 2008; Wittwer, Nückles, & Renkl, 2010). While this adaptive on-demand design may be effective, the technological resources it requires currently limit its wide application. Determining what information to provide for the learner and what to require them to generate, such as with instructional- and self-explanations, is a fundamental open problem in instructional science (Koedinger & Aleven, 2007). The current study presents a method of integrating self- and instructional-explanations that may be able to provide some resolution to these mixed findings.

The generation and correction process seems to harnesses the benefits of both types of explanations. In this process, learners generate a self-explanation, read an instructional explanation, and then correction their own explanation. Prompted self-explanations often are not of high quality, indicating that it is difficult for learners to optimally engage in self-explanation (Chi et al., 1989; McEldoon et al., 2012; Renkl, 2002; Roy & Chi, 2005). This limitation can be overcome when self-explanations are followed by instructional explanations. In a one-on-one tutoring contexts, the sequencing of explanations supports reflection and comparison (Roy & Chi, 2005). Building an explicit correction process into the learning activity can foster the noticing of discrepancies between the learner's own possibly incomplete or incorrect self-explanation and the correct instructional explanation. Once a discrepancy is noticed, the learner can repair their original representation, resulting in learning (Chi, 2000). Noticing these points of mismatch have been shown to be especially effective for driving learning (Van Lehn, 1998). Overall, a correction process integrates the correct instructional explanation into the learner's ongoing knowledge construction activities (Berthold & Renkl, 2010). This integration is especially important, as simply receiving correct information does not ensure understanding. The current study provides evidence that a generation and correction process, when enacted appropriately, may support greater learning.

Focus on Goal-Operator Combinations

Another way a learner can increase their focused processing is through an understanding the *goal* of the problem-solving step, and how the values and *operators* within it achieve this *goal*. My focused processing stance specifies not only that the learner should focus on procedures, but specifically that they should focus on goal-operator combinations. One way to achieve this is to include gaps within a worked example that require the learner to assign meaning to values within the problem-solving steps. This study contrasted incomplete worked examples with gaps that do and do not focus the learner on goal-operator combinations. When students completed learning activities that included completing gaps that focus on goal-operators, there was a week one benefit relative to other learning activities that included processing the same worked example in a less focused way. This finding aligns with other recent findings on incomplete worked examples, although this study is the first to the author's knowledge that directly tests for the benefit of gaps that focus on goal-operator combinations.

The current study directly tested the benefit of the nature of incomplete worked examples. Both conditions had gaps-to-be-filled, but tested the effects of *what* should be incomplete. Students either filled in gaps that facilitated the integration of operators and goals and required a search across multiple sources of information within the worked example or filled in gaps that could be completed by merely attending to the immediate step itself. This study provides evidence that increasing the focused processing of a problem solving procedure through scaffolding the learners' attention on goal-operator combinations can be an effective way to increase knowledge transfer. That a learner “not only knows the procedural steps for problem-solving tasks, but also understands when to deploy them and why they work” (Gott, Glaser, Hall, Dibble, & Pokorny, 1996), is considered essential. This ability to recognize and flexibly apply the relevant parts of a previously learned procedure to solve novel problems (i.e. *transfer*) is an important goal of learning (e.g. Catrambone, 1996; Gott et al., 1996; Paas & van Gog, 2006).

There are several ways in which a focus on goal-operator combinations could increase learning. Incomplete worked examples that direct the learners’ attention to goal-operator combinations can be thought of as modeling an expert's attentional focus. It is well demonstrated that experts focus on deep relational features of a problem, whereas novices tend of focus on surface features (Chase & Simon, 1973; Chi, Feltovich, & Glaser, 1981; Gentner & Toupin, 1986; Kotovsky & Gentner, 1996; Larkin, 1983; Simon & Simon, 1978; Sloutsky & Yarlas, 2000; Yarlas & Sloutsky, 1999). Filling in intermediate gaps that require a consideration of goal-operator combinations can explicitly direct the novices' attention to aspects of the problem that an expert would process and encode. This process could also direct them away from dwelling on less important surface features.

Another way a focus on goal-operator combinations is thought to increase the quality of focused processing is through integrating information. The process of assigning meaning to values often requires mapping and integrating between multiple sources of information (Renkl, 2011; Seufert & Brünken, 2004). For example, in order to determine a goal-operator combination such as *‘You can calculate this arc of a circle [goal] by subtracting 33" from 360" [operators]’*, a hypothetical learner might have to interpret and then map together information from a diagram of a circle, a geometry theorem, the problem statement with the particular values, and perhaps some instructional explanations about calculating arcs. A full understanding of the relationship or *link* between goals and operations often requires drawing information together from disparate sources. Other ways this integration has been facilitated is through the use of animated agents who use gaze and gesture to direct the learners' attention, and color coding and flashing within an interactive learning environment (Atkinson, 2002; Berthold & Renkl, 2009). In contrast, when incomplete worked examples do not sufficiently scaffold the assignment of meaning to the problem-solving steps, they are not as effective (e.g. Schwonke et al., 2011). A learner's ability to articulate such links when learning from worked examples has been correlated with both procedural and conceptual knowledge and ability to solve related but novel problems (Schwonke et al., 2009). Therefore, well-designed worked examples that focus on goal-operator combinations may facilitate learning when they require the learner to actively link the operators within a problem solving procedure with the goal it is intended to accomplish by integrating information.

Overall, a deeper understanding of the learning domain can be enhanced when the learner is scaffolded to linking the operators with the respective goals, and thereby developing their current understanding of the domains concepts and procedures.

In sum, this study supports my modified focused processing stance that more clearly specifies two important aspects learners should process in order to achieve learning gains.

Alignment with Other Learning Frameworks

The findings of this study also supports hypotheses put forth in two other areas of cognition research. Generally, the finding that more focused processing supports greater learning is aligned with the levels of processing framework from the memory literature. The depth of processing a stimulus undergoes during learning has been shown to correspond with the strength of its later memory trace, or how well it is learned (Craik & Lockhart, 1972; Craik, 2002). "Deeper" refers to the analysis of meaning, inference, and implication. For example, when shown a list of words, participants who were asked to judge if the word starts with a capital letter (i.e., shallow processing) did not remember the word list after a delay as well as those who made a semantic judgment about the word (e.g. Is this an animal?; deeper processing). A recent meta-analysis initially identified 221 studies that dealt with levels of processing, and 7 that met strict inclusion criteria (e.g. measures of processing during task performance and a distinction between processing and performance) did indeed demonstrate a relationship between depth of processing and learning outcomes (average effect size, $b = 0.25$; median = 0.35) (Dinsmore & Alexander, 2012). As "deeper" refers to the analysis of meaning, inference, and implication, the moreFP condition engaged in deeper processing than the lessFP condition (Craik, 2002).

Echos of this general cognitive learning principle can be heard in contemporary educational and psychological thinking. Engaging learners in *actively or deeply* processing and constructing their own knowledge or mental models has long been put forth as an effective way of inducing learning (Bruner, 1961; Mayer, 2009; Piaget, 1970). There are many constructivist theories of learning (e.g. Piaget, 1964; Steffe & Kieren, 1994, etc.), but the particular cognitive perspective put forth by Mayer (2009) is most relevant. In this view, constructivism is a theory of learning in which the learner builds knowledge structures in working memory by engaging in *active cognitive processing* during learning (Bransford, Brown, & Cocking, 1999; Chi, 2009; Mayer, 2009).

Recently, the general idea of cognitive constructivism has been formulated into a hierarchy of learning activities by Chi (2009). She posits four hierarchical levels of processing activities in instructional contexts; passive, active, constructive, and interactive. Each type of learning activity is predicted to result in greater learning gains than the one before it. Active learning requires some selection process or physical activity while learning, such as underlining, pointing, copying problem steps, or manipulating or selecting aspects of the problem. The cognitive processes involved in active activities may include activating, assimilating, encoding, storing, or searching existing knowledge. The idea is that if the learner is engaged in the specified overt activities, presumably the linked cognitive processes are also taking place. Contrast this with constructive learning activities, which require that the learner is generative and produces outputs that contain ideas that go beyond the information presented. Some constructive activities could be explaining, justifying, connecting, reflecting, planning or predicting. These activities are thought to utilize the cognitive processes of inferring new knowledge, integrating new and excising information, organizing knowledge for coherence, or repairing faulty knowledge. There is much evidence to support the idea that the more constructively the learner engages with the material and builds their own knowledge, the greater their learning is (see Fonseca & Chi, 2009 for a review).

The lessFP and moreFP conditions are representative of active and constructive levels, respectively. The lessFP condition engaged in *copying* instructional explanations and *manipulating* aspects on the worked example problem solving steps by calculating the final value. The moreFP condition engaged in *explanation generation* and worked example completion that required *generating connections*. Indeed, the auxiliary survey included self-report items that tapped student engagement in these specific processes. Participants in both conditions reported equal levels of 'active' activities, whereas the moreFP condition reported higher levels of 'constructive' activities. The findings of this study support this framework's claim that engaging in activities that can be considered constructive can result in greater learning gains than those that are considered active. Past research has often contrasted the benefit of incomplete worked examples or self-explanations against a 'passive' activity, according to this framework. The current study used a more rigorous control group.

The results of this study converge with prior work that has also found a benefit of learning activities that may be considered 'constructive' over those that are 'active'. For example, following Chi's (2009) operationalizations of these levels, the constructive activities of taking notes (Trafton & Trickett, 2001), asking questions (Graesser & Person, 1994), posing problems (Mestre, 2001), comparing and contrasting cases (Schwartz & Bransford, 1998), and generating predictions (Klahr & Nigam, 2004), have all been shown to be more beneficial for learning than active control tasks.

Absence of Effect at Week Two

The lack of effect at week two provides a cautionary tale. The pattern of results at week two may be due to a few potential reasons. One reason may have been reduction in student engagement with the learning materials. An important aspect of effective instructional designs is how appropriately the learners engage with them. There is clear evidence of lower levels of student instructional compliance in week two relative to week one in the current study. In week one, the moreFP condition correctly completed more of the worked example than the less FP condition. Across weeks, the moreFP condition had significantly higher original part one explanation quality as well as higher correct worked example accuracy in week one than they did in week two. Levels of compliance were the same in both conditions in week two. Also, the moreFP participants' intervention compliance in week one was greater than in week two, where there were no differences compared to the lessFP condition in week two. The data suggest that the moreFP participants put more effort into their initial explanation generation and worked example accuracy in week one compared to the lessFP students at week one. Therefore, the moreFP students's greater intervention compliance at week one may be a contributing factor to the resulting learning gains. This could be a potential reason for the benefit of the moreFP materials on learning outcomes in week one, but not at week two.

The quality of the participants' work dropped from week one to week two. This could be due to either the week two work being more challenging, or participant effort being lower. If the materials in week two were simply more difficult, one would expect a higher number of corrections to their part one worked examples once they received the correct values in part two. However, this was not the case. Instead, the number of corrections was constant between weeks. It seems plausible that reduced performance in week two could be due to time constraints, but self-reported levels of being complete and wanting more time were the same as week one. Taken together, this evidence suggests that the students did not put in as much effort into the week two activities as they did in week one. Therefore, lower levels student effort may have been a contributing factor to the lack of effect in week two.

Anecdotal evidence support this conclusion. At week two, a few participants reported that they knew the answers would be coming in the part two packet, so did not try as hard to complete the worked example and explanation prompts on part one as they did in week one. Also, many participants asked if their work would be graded, and when they were told it would not be, they did not seem to put in as much effort into their work.

Other differences that could possibly account for the lack of condition differences in week two. The learning activity and assessments may not have been well designed. For example, all the knowledge of concepts items had to be dropped on the week two assessment due to poor psychometric qualities. This likely lessened the sensitivity of the week two measures. Future research should use better developed and validated assessments.

Another potential explanation is that by week two, the learners were no longer sufficiently novice for the condition manipulation to have as strong of a benefit. Recall that the topic for week one was one-way ANOVA, and for week two it was a two-way ANOVA. These topics share many of the same foundational concepts and computational procedures. At week one, the participant's pretest knowledge of ANOVA was low (score average 46%). However, the week two pretest scores on two-way ANOVA were significantly higher than week one (average score 59%). It is possible that this level of prior knowledge lessened the impact of the condition manipulation. Recall that worked examples are particularly effective for novices. Indeed, worked examples lose their effectiveness once a learner gains a sufficient amount of domain knowledge (Kalyuga et al., 2003, 2000, 2001). The scaffolding and support that a worked example provides is no longer necessary and become redundant and ineffective. It is plausible that the students were no longer sufficiently novice to benefit as much from the extra supports in the more focused processing condition.

Limitations & Future Directions

The primary finding of this study is that if learners engage in more focused processing, learning outcomes can increase. The key to this is ensuring that the learners engage appropriately. The learners in the study did so at week one, but week two compliance levels were significantly lower. Fortunately, one straightforward way to remedy this limitation in the future is to hold students accountable for the more focused processing activities by assigning grades to their initial work, or to provide some other incentive. Another factor that may have contributed to the lower levels of compliance in week two is lack of time. Many students reported wanting more time to complete the activities across both weeks. Allowing ample time for the participants to engage with the learning materials may result in stronger effects. Another possibility is that the intervention activity in week two may not have been as well designed, and thus did not foster as much learning as the week one activity.

The current study was completed within the context of learning statistics at a university with high-achieving students. The generalization of these findings to other domains and other students groups should be investigated further.

Future studies should be designed to specifically test for knowledge of goal-operator combinations as well as of specific principles. An understanding of these combinations and principles were built into the current assessments, but more insight could be gained if these specific types of knowledge could be measured and analyzed independently.

Similarly, the benefits of incomplete worked examples that focus on goal-operator combinations and focused processing of principles through explanation generation and correction are currently conflated in the current study. Future studies should examine the role of these focused processing methods independently, and determine the nature of the relationship between these two factors. Additionally, there may be other and possibly more effective ways to foster focused processing. This study provides some support for the importance of focused processing on both concepts and procedures via explanation generation and correction and incomplete worked examples that highlight goal-operator combinations. Further refinement of the focused processing stance should continue to expand instructional activities that support these processes.

REFERNCES

Aleven, V. A. W. M. M., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147–179. doi:10.1207/s15516709cog2602

Anderson, J R, Fincham, J. M., & Douglass, S. (1997). The role of examples and rules in the acquisition of a cognitive skill. *Journal of experimental psychology. Learning, memory, and cognition*, *23*(4), 932–45. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9231437

Anderson, John R. (1993). Rules of the Mind 1- Production Systems and the ACT-R Theory 2- Knowledge Representation. In *Rules of the Mind*. Hillsdale, NJ: Erlbaum.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*(2), 416–427. doi:10.1037//0022-0663.94.2.416

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, *70*(2), 181–214. doi:10.3102/00346543070002181

Atkinson, R. K., & Renkl, A. (2007). Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples. *Educational Psychology Review*, 375–386. doi:10.1007/s10648-007-9055-2

Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: effects of self-explanation prompts and fading worked-out steps. *Journal of Educational Psychology*, *95*(4), 774–783. doi:10.1037/0022-0663.95.4.774

Berthold, K., Eysink, T. H. S., & Renkl, A. (2009). Assisting self-explanation prompts are more effective than open prompts when learning with multiple representations. *Instructional Science*, *37*(4), 345–363. doi:10.1007/s11251-008-9051-z

Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, *101*(1), 70–87. doi:10.1037/a0013247

Berthold, K., & Renkl, A. (2010). How to foster active processing of explanations in instructional communication. *Educational Psychology Review*, *22*(1), 25–40. doi:10.1007/s10648-010-9124-9

Berthold, K., Röder, H., Knörzer, D., Kessler, W., & Renkl, A. (2011). The double-edged effects of explanation prompts. *Computers in Human Behavior*, *27*(1), 69–75. doi:10.1016/j.chb.2010.05.025

Bertsch, S., Pesta, B. J., Wiscott, R., & McDaniel, M. A. (2007). The generation effect: a meta-analytic review. *Memory & cognition*, *35*(2), 201–10. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17645161

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition and Instruction*, *13*(2), 221–252. doi:10.1207/s1532690xci1302

Bisanz, J., & LeFevre, J.-A. (1992). Understanding elementary mathematics. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skill* (pp. 113–136). Amsterdamn: Elsevier Science.

Bransford, J. D., Brown, A., & Cocking, R. (1999). *How people learn: Mind, brain, experience, and school*. Washington, DC: National Research Council.

Bruner, J. S. (1961). The act of discovery. *Harvard Educational Review*, *31*, 21–32.

Bujak, K. R. (2010). *A framework of passive-active-constructive study techniques: A divergence between assigned and reported behaviors*. Georgia Institute of Technology.

Calin-Jageman, R. J., & Ratner, H. H. (2005). The role of encoding in the self-explanation effect. *Cognition and Instruction*, *23*(906608910), 523–543. doi:10.1207/s1532690xci2304

Carroll, W. M. (1994). Using worked examples as an instructional support in the algebra classroom, *86*(3), 360–367.

Catrambone, R. (1996). Generalizing solution procedures learned from examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(4), 1020–1031.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, *127*(4), 355–376. doi:10.1037/0096-3445.127.4.355

Catrambone, R., & Yuasa, M. (2006). Acquisition of procedures: The effects of example elaborations and active learning exercises. *Learning and Instruction*, *16*(2), 139–153. doi:10.1016/j.learninstruc.2006.02.002

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Chi, M. T. H. (2000). Self-Explaining Expository Texts: The Dual Processes of Generating Inferences and Repairing Mental Models. In R. Glaser (Ed.), *Advances in Instructional Psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, *1*(1), 73–105. doi:10.1111/j.1756-8765.2008.01005.x

Chi, M. T. H., Bassok, M. M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explantions: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182. doi:10.1207/s15516709cog1302_1

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, *5*(2), 121–152. doi:10.1207/s15516709cog0502_2

Chi, M. T. H., Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, *18*(3), 439–477. doi:10.1207/s15516709cog1803_3

Conati, C., & Van Lehn, K. (1999). Teaching meta-cognitive skills: Implementation and evaluation of a tutoring system to guide self-explanation while learning from examples. *Cognitive Science*, 1–8.

Conati, C., & Van Lehn, K. (2000). Toward computer-based support of meta-cognitive skills: A computational framework to coach self-explanation. *International Journal of Artificial Intelligence in Education*, *11*, 389–415.

Craik, F. I. M. (2002). Levels of processing: past, present. and future? *Memory*, *10*(5-6), 305–18. doi:10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. doi:10.1016/S0022-5371(72)80001-X

DeCaro, M. S., & Rittle-Johnson, B. (2012). Exploring mathematics problems prepares children to learn from instruction. *Journal of experimental child psychology*, *113*(4), 552–68. doi:10.1016/j.jecp.2012.06.009

Dinsmore, D. L., & Alexander, P. A. (2012). A Critical Discussion of Deep and Surface Processing: What It Means, How It Is Measured, the Role of Context, and Model Specification. *Educational Psychology Review*. doi:10.1007/s10648-012-9198-7

Duffy, G. G., Roehler, L. R., Meloth, M. S., & Vavrus, L. G. (1986). No Title. *Teaching and Teacher Education*, *2*, 197–214. doi:http://dx.doi.org/10.1016/S0742-051X(86)80002-6

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*(3), 83–87. doi:10.1111/1467-8721.01235

Fonseca, B. A. B. A., & Chi, M. T. H. (2011). Instruction based on self-explanation. In R. Mayer & P. Alexander (Eds.), *The Handbook of Research on Learning and Instruction* (pp. 296–321). New York, NY: Routledge Press.

Gelman, A. (2005). Analysis of variance: Why it is more important than ever. *The Annals of Statistics*, *33*(1), 1–53. doi:10.1214/009053604000001048

Gentner, D., & Toupin, C. (1986). Systematicity and Surface Similarity in the Development of Analogy, *277300*.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction*, *16*(2), 104–121. doi:10.1016/j.learninstruc.2006.02.007

Gott, S. P., Glaser, R., Hall, E. P., Dibble, E., & Pokorny, R. A. (1996). A naturalistic study of transfer. In *Adaptive expertise in technical domains* (pp. 258–288). Ablex.

Graesser, A. C., & Person, N. K. (1994). Question-asking during tutoring. *American Educational Research Journal*, (31), 104–137.

Greeno, J. G., Riley, M. S., & Gelman, R. (1984). Conceptual competence and children's counting. *Cognitive Psychology*, *16*(1), 94–143. doi:10.1016/0010-0285(84)90005-7

Große, C. S., & Renkl, A. (2004). Learning from Worked Examples: What Happens if Errors are Included? In P. Gerjets, J. Elen, R. Joiner, & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning* (pp. 356–364). Tuebingen: Knowledge Media Research Center.

Große, C. S., & Renkl, A. (2006a). Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, *16*(2), 122–138. doi:10.1016/j.learninstruc.2006.02.001

Große, C. S., & Renkl, A. (2006b). Effects of multiple solution methods in mathematics learning. *Learning and Instruction*, *16*(2), 122–138. doi:10.1016/j.learninstruc.2006.02.001

Hausmann, R. G. M., & Van Lehn, K. (2007). Explaining self-explaining: A contrast between content and generation. In *Proceeding of the 2007 conference on Artificial Intelligence in Education*. Amsterdam, The Netherlands: IOS Press.

Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, *18*(1), 54–65. doi:10.1016/j.learninstruc.2006.10.008

Hilbert, T. S., Schworm, S., & Renkl, A. (2004). Learning from worked-out examples: The transition from instructional explanations to self-explanation prompts. In P. Gerjets, J. Elen, R. Joiner, & P. Kirschner (Eds.), *Instructional design for effective and enjoyable computer-supported learning* (pp. 184–192). Tubingen, Germany: Knowledge Media Research Center.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The Expertise Reversal Effect. *Educational Psychologist*, *38*, 23–31. doi:10.1207/S15326985EP3801_4

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126–136. doi:10.1037//0022-0663.92.1.126

Kalyuga, S., Chandler, P., Tuovinen, J., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, *93*(3), 579–588. doi:10.1037//0022-0663.93.3.579

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*, 75–86.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, *15*(10), 661–667.

Koedinger, K. R., & Aleven, V. A. W. M. M. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, *19*(3), 239–264. doi:10.1007/s10648-007-9049-0

Kotovsky, L., & Gentner, D. (1996). Comparison and Categorization in the Development of Relational Similarity. *Child Development*, *67*(6), 2797–2822.

Kuhn, D., & Katz, J. (2009). Are self-explanations always beneficial? *Journal of experimental child psychology*, *103*(3), 386–94. doi:10.1016/j.jecp.2009.03.003

Landis, R., & Koch, G. (1977). The measurement of observer agreement for catagorical data. *Biometrics*, *33*(1), 159–174. doi:10.2307/2529310

Larkin, J. (1983). The Role of Problem Representation in Physics. In *Mental Models* (p. 75).

LeFevre, J.-A., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction*, *3*, 1–30.

Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. *Handbook of research on teaching*, 333–357.

Leinhardt, G., & Steele, M. D. (2005). Seeing the complexity of standing to the side: Instructional dialogues. *Cognition and Instruction*, *23*, 87–163.

Matthews, P. G., & Rittle-Johnson, B. (2009). In pursuit of knowledge: Comparing self-explanations, concepts, and procedures as pedagogical tools. *Journal of experimental child psychology*, *104*(1), 1–21. doi:10.1016/j.jecp.2008.08.004

Mayer, R. E. (2009). Constructivism as a theory of leaning versus constructivism as a prescription for instruction. In S. Tobias & T. M. Duffy (Eds.), *Constructivist Instruction: Success or Failure?* (pp. 184–200). New York: Routledge Press.

McEldoon, K.L., Durkin, K. L., Rittle-Johnson, B. (2012). Is Self-Explanation Worth the Time? A Comparison To Additional-Practice. British Journal of Educational Psychology. DOI: 10.1111/j.2044-8279.2012.02083.x

Merrill, D. C., Reiser, B. J., Merrill, S. K., & Landes, S. (1995). Tutoring: Guided learning by doing. *Cognition and Instruction*, 315–372.

Mestre, J. P. (2001). Implications of research on learning for the education of prospective science and physics teachers. *Physics Education*, *36*, 44–51.

Moreno, R. (2006). When worked examples don't work: Is cognitive load theory at an Impasse? *Learning and Instruction*, *16*(2), 170–181. doi:10.1016/j.learninstruc.2006.02.006

Mwangi, W., Sweller, J. (1998). Learning to solve compare word problems: The effect of example format and generating self-explanations. *Cognition and Instruction*, *16*(2), 173–199. doi:10.1207/s1532690xci1602

Otieno, C., Freiburg, D.-, Schwonke, R., & Renkl, A. (2011). Measuring learning progress via self-explanations versus problem solving: A suggestion for optimizing adaptation in intelligent tutoring systems. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 84–89). Boston, MA: Cognitive Science Society.

Paas, F. G. W. C. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paas, F. G. W. C., & van Gog, T. (2006). Optimising worked example instruction: Different ways to increase germane cognitive load. *Learning and Instruction*, *16*(2), 87–91. doi:10.1016/j.learninstruc.2006.02.004

Paas, F. G. W. C., & van Merriënboer, J. J. G. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, *86*(1), 122–133. doi:10.1037/0022-0663.86.1.122

Piaget, J. (1970). *Science of education and psychology of the child*. New York: Oxford University Press.

Piaget, Jean. (1964). Development and learning. In R. E. Ripple & V. N. Rockcastle (Eds.), *Piaget Rediscovered* (1997th ed., pp. 7–20). New York: Company, W. H. Freeman and.

Pirolli, P. L., & Anderson, J. R. (1985). The role of learning from examples in the acquisition of recursive programming skills. *Canadian Journal of Psychology*, *39*(2), 240–272.

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144–161.

Recker, M., & Pirolli, P. L. (1995). Modeling individual differences in students' learning strategies. *The Journal of the Learning Sciences*, *4*, 1–38.

Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction*, *16*(2), 92–103. doi:10.1016/j.learninstruc.2006.02.008

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science*, *21*(1), 1–29. doi:10.1207/s15516709cog2101_1

Renkl, A. (1999). Learning mathematics from worked-out examples: Analyzing and fostering self-explanations. *European Journal of Psychology of Education*, *14*, 477–488. doi:10.1007/BF03172974

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556. doi:10.1016/S0959-4752(01)00030-5

Renkl, A. (2011). Instruction Based on Examples. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of Research on Learning and Instruction* (pp. 272–295). New York, NY: Routledge Press.

Renkl, A., & Atkinson, R. K. (2007). Interactive learning environments: Contemporary issues and trends. An introduction to the special issue. *Educational Psychology Review*, *19*(3), 235–238. doi:10.1007/s10648-007-9052-5

Renkl, A., Atkinson, R. K., & Große, C. S. (2004). How fading worked solution steps works: A cognitive load perspective. *Instructional Science*, *32*(1/2), 59–82. doi:10.1023/B:TRUC.0000021815.74806.f6

Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From example study to problem solving: Smooth transitions help learning. *The Journal of Experimental Education*, *70*(4), 293–315. doi:10.1080/00220970209599510

Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from worked-out examples: The effects of example variability and elicited self-explanations. *Contemporary Educational Psychology*, *23*(1), 90–108. doi:10.1006/ceps.1997.0959

Rittle-Johnson, B. (2006). Promoting transfer: effects of self-explanation and direct instruction. *Child development*, *77*(1), 1–15. doi:10.1111/j.1467-8624.2006.00852.x

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of Educational Psychology*, *91*(1), 175–189. doi:10.1037/0022-0663.91.1.175

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of Educational Psychology*, *93*(2), 346–362. doi:10.1037//0022-0663.93.2.346

Robins, S., & Mayer, R. E. (1993). Schema training in analogical reasoning. *Journal of Educational Psychology*, *85*(3), 529–538.

Ross, B. H. (1989). Distinguishing types of superficial similarities: Different effects on the access and use of earlier problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 456–468. doi:10.1037/0278-7393.15.3.456

Roy, M., & Chi, M. T. H. (2005). The self-explanation principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 271–286). Cambridge, UK: Cambridge University Press.

Sánchez, E., García-Rodicio, H., & Acuña, S. R. (2008). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, *37*(6), 537–563. doi:10.1007/s11251-008-9074-5

Schnotz, W., & Kürschner, C. (2007). A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, *19*(4), 469–508. doi:10.1007/s10648-007-9053-4

Schwartz, D., & Bransford, J. (1998). A Time For Telling. *Cognition and Instruction*, *16*(4), 475–5223. doi:10.1207/s1532690xci1604_4

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology*, *103*(4), 759–775. doi:10.1037/a0025140

Schwartz, D. L., & Martin, T. (2004). Inventing to Prepare for Future Learning : in Statistics Instruction o t o y p o o t. *Cognition*, *22*(2), 129–184.

Schwonke, R., Renkl, A., Krieg, C., Wittwer, J., Aleven, V. A. W. M. M., & Salden, R. (2009). The worked-example effect: Not an artefact of lousy control conditions. *Computers in Human Behavior*, *25*(2), 258–266. doi:10.1016/j.chb.2008.12.011

Schwonke, R., Renkl, A., Salden, R., & Aleven, V. A. W. M. M. (2011). Effects of different ratios of worked solution steps and problem solving opportunities on cognitive load and learning outcomes. *Computers in Human Behavior*, *27*(1), 58–62. doi:10.1016/j.chb.2010.03.037

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education*, *46*(4), 426–445. doi:10.1016/j.compedu.2004.08.011

Scribner, S. (1984). Studying working intelligence. In M. Rogoff & J. Lave (Eds.), *Everyday cognition: Its development in social context* (pp. 9–28). Cambridge, MA: Harvard University Press.

Seufert, T., & Brünken, R. (2004). Supporting coherence formation in multimedia learning. In *Proceedings of the first joint meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers* (pp. 138–147).

Siegler, R. S., & Lin, X. (2010). Self-explanations promote children's learning. In H. S. Waters & W. Schneider (Eds.), *Metacognition, Strategy Use, and Instruction* (pp. 85–112).

Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. *Children's thinking: What develops*, 325–348.

Sloutsky, V. M., & Yarlas, A. S. (2000). Problem Representation in Experts and Novices : Part 2 . Underlying Processing Mechanisms. *Cognitive Science*.

Stark, R. (1999). *Lernen mit Losungsbeispielen. Der Einfluß unvollstandiger Losungsschritte auf Beispie- lelaboration. Motivation und Lernerfolg [Learning by worked-out examples. The impact of incomplete solution steps on example elaboration, motivation and learning outcom*. Bern, CH: Huber.

Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction*, *12*(1), 39–60. doi:10.1016/S0959-4752(01)00015-9

Steffe, L. P., & Kieren, T. (1994). Radical constructivism and mathematics education. *Journal for Research in Mathematics Education*, *25*(6), 711–733.

Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, *12*(2), 257–285. doi:10.1207/s15516709cog1202_4

Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, *16*(2), 165–169. doi:10.1016/j.learninstruc.2006.02.005

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, *2*(1), 59–89.

Sweller, J., Merrienboer, J. J. G. Van, & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology*, *10*(3), 251–296.

Tarmizi, R. a., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology*, *80*(4), 424–436. doi:10.1037/0022-0663.80.4.424

Trafton, J. G., & Trickett, S. B. (2001). Note-taking for self-explanation and problem solving. *Assessment*, *16*, 1–38.

Treagust, D. F., & Harrison, A. G. (1999). The genesis of effective scientific explanations for the classroom. *Researching teaching: Methodologies and practices for understanding pedagogy*, 28–43.

Van Gog, T., Paas, F. G. W. C., & van Merriënboer, J. J. G. (2006). Effects of process-oriented worked examples on troubleshooting transfer performance. *Learning and Instruction*, *16*(2), 154–164. doi:10.1016/j.learninstruc.2006.02.003

Van Lehn, K. (1998). Analogy events: How examples are used during problem solving. *Cognitive Science*, *22*, 347–388.

Van Lehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *21*, 209–249. doi:10.1207/S1532690XCI2103

Van Merriënboer, J. J. G. (1990). Strategies for programming instruction in high school: Program completion vs. program generation. *Journal of educational computing research*, *6*, 265–285.

Van Merriënboer, J. J. G., & De Croock, M. B. (1992). Strategies for computer-based programming instruction: Program completion vs. program generation. *Journal of Educational Computing Research*, *8*, 365–394.

Ward, M., & Sweller, J. (1990). Structuring effective worked examples. *Cognition and Instruction*, *7*(1), 1–39.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Educational Research*, *13*, 21–39.

Webb, N. M., & Farivar, S. (1999). Developing productive group interaction in middle school mathematics. In A. M. O'Donnell & A. King (Eds.), *Cognitive perspectives on peer learning* (pp. 117–149). Mahwah: Erlbaum.

Wittwer, J., Nückles, M., & Renkl, A. (2010). Using a diagnosis-based approach to individualize instructional explanations in computer-mediated communication. *Educational Psychology Review*, *22*(1), 9–23. doi:10.1007/s10648-010-9118-7

Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, *43*(1), 49–64. doi:10.1080/00461520701756420

Wittwer, J., & Renkl, A. (2010). How effective are instructional explanations in example-based learning? A meta-analytic review. *Educational Psychology Review*, *22*(4), 393–409. doi:10.1007/s10648-010-9136-5

Wong, R. M., Lawson, M. J., & Keeves, J. (2002). The effects of self-explanation training on students' problem solving in high-school mathematics. *Learning and Instruction*, *12*, 233–262.

Yarlas, A. S., & Sloutsky, V. M. (1999). Representation Of Arithmetic Principles By Novices: Knowledge Or Attention?

APPENDIX A

Name: ____

Section Time: **3 minutes**

Keep your answers brief

Wait until your TA says to begin

Start Time: ______________

Section Time: 5 minutes

1) At a high level, how does an ANOVA determine if there are differences between groups?

2) Briefly list the steps needed calculate an ANOVA.

3) How do the steps involved in a one-way ANOVA answer the question we are trying to test?

Stop Time: ______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Two

For each section:

1) In column 1, read through the definitions carefully, and fill in the blanks with the appropriate abbreviation
2) In column 2, complete the calculations
3) Read the thought questions carefully
4) Fill in the Summary Table as you solve each part

See your TA for Part Three

Start Time: ____________

One-Way ANOVA

The Dataset

	Group 1	Group 2	Group 3
	3	8	21
	5	6	16
	1	11	32
	0	7	16
	7	4	25
	2	12	40
$\bar{X}$	3	8	25

Fill in the Summary Table

Source	*df*	*SS*	*MS*	*F*
Across Group				
Within Group				
Total				

Ho : $\mu_{Group1} = \mu_{Group2} = \mu_{Group3}$

SUMS OF SQUARES

Definitions	Calculations
x_{ij} = *individual score*; $\bar{x}_j$ = *group mean*; $\bar{x}_{..}$ = *grand mean* The Sum of Squares **total** (____) is the sum of the squared deviations of all the **individual scores** from the **grand (or overall) mean**. *This is a measure of the total amount of variability in the data.* $SS_{TOT} = SS_{AC} + SS_{WG}$	$SS_{TOT} = \Sigma(x_{ij} - \bar{x}_{..})^2$ $(3-12)^2+(5-12)^2+(1-12)^2+(0-12)^2+(7-12)^2+(2-12)^2+(8-12)^2+(6-12)^2+(11-12)^2+(7-12)^2+(4-12)^2+(12-12)^2+(21-12)^2+(16-12)^2+(32-12)^2+(16-12)^2+(25-12)^2+(40-12)^2$ = ________
The Sum of Squares **across groups** (____) is the sum of the squared deviations of **each group mean** from the **grand (or overall) mean**, times the number of datapoints in that group. We multiply by the number of scores in the group to get a weighted estimate of the variability. *This is a measure of the amount of variability due to the treatment effect and due to noise.* SS_{AG} = *treatment effect + noise*	
The Sum of Squares **within groups** (____) is the sum of the squared deviations of the **individual scores** from **their group mean**. *This is a measure of the amount of variability within each group due to noise.* SS_{WG} = *noise*	$SS_{WG} = \Sigma(\bar{x}_{ij} - \bar{x}_j)^2$ $(3-3)^2+(5-3)^2+(1-3)^2+(0-3)^2+(7-3)^2+(2-3)^2+(8-8)^2+(6-8)^2+(11-8)^2+(7-8)^2+(4-8)^2+(12-8)^2+(21-25)^2+(16-25)^2+(32-25)^2+(16-25)^2+(25-25)^2+(40-25)^2$ = ________

Thought Questions

- **What main idea are these SSs capturing?**

The main idea is that we are capturing the *amount of variability* in the data. SS_{TOT} captures all of it, and SS_{AG} and SS_{WG} partition this variability into parts that are due to the treatment effect and noise.

- **Why would you expect SS_{AG} to be *different* when H_O is true rather than false? Consider whether you would expect SS_{AG} to be large or small when H_O is true and when H_O is false, explaining your reasoning.**

The value of SS_{AG} would be different if H_O were true or false, because SS_{AG} captures how different each group mean is from the grand mean. If H_O were true, the group means would be very similar to the grand mean, resulting in small differences, and a small value for the SS_{AG}. If H_O were false, the group means would be different from the grand mean, resulting in larger

differences, and a larger value for the SS_{AG}.

MEAN SQUARES & F-RATIO	
Definitions	**Calculations**
k = Number of Groups; N = total number of datapoints The Mean Square (____) is the amount of variance within each partition. MS has the same formula as variance (S^2): $\mathbf{MS} = \frac{SS}{df} = \frac{SS}{n-1} = S^2$	$\mathbf{MS}_{AG} = \frac{SS_{AG}}{df_{AG}} = \frac{1596}{2} =$ _______ $\mathbf{df}_{AG} = (k-1) = (3-1) =$ _____
It can be thought of as the *average* (or mean) variability per degree of freedom.	$\mathbf{MS}_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{532}{15} =$ _______ $\mathbf{df}_{WG} = df_{TOT} - df_{AG} = (N-1)-(k-1)$ $= (18-1)-(3-1) = (17-2) =$ _____
The F-ratio puts the amount of variability among the groups over the amount of variability within each group.	$\mathbf{F}_{\_\_,\_\_} = \frac{MS_{AG}}{MS_{WG}} = \frac{798}{35.5} =$ _______

Thought Questions

- **If H_O is true, is MS_{AG} smaller or larger than MS_{WG}? If H_O is false? Explain how you know.**

If H_O is true, MS_{AG} and MS_{WG} should be approximately equal. MS_{AG} includes variability due to the treatment effect and due to noise. Since there is no treatment effect, MS_{AG} would only include variability due to noise. MS_{WG} only includes variability due to noise anyway, so they should be about the same.
If H_O is false, MS_{AG} should be larger than the MS_{WG}. MS_{AG} includes variability due to the treatment effect *and* due to noise. Since there is a treatment effect, MS_{AG} would include this variability *plus* the variability due to noise. MS_{WG} only includes variability due to noise. So, MS_{AG} should be larger.

- **How does the F-ratio quantify the *treatment effect*?**

The F-ratio divides the MS_{AG} by the MS_{WG}. The MS_{AG} contains variability due to the treatment effect *and* noise, and dividing it by the MS_{WG}, which only contains variably due to noise, will cancel out the variability due to noise. The resulting value of the F-ratio reflects the size of variability due to the treatment effect.

STATISTICAL CONCLUSION	
Definitions	**Calculations**
This is the critical value by which we judge if the amount of variability between groups is less than 5% likely to occur between groups from the same population.	$_{.05}\mathbf{F}_{\_\_,\_\_} = 3.68$ *(from table)* Decision: **<u>Reject</u>** Fail to Reject

Thought Question

- **Interpret your conclusion in terms of the treatment effect. Discuss variability *among* groups and *within* groups.**

There is a treatment effect, and not all of the group means are the same. The variance among groups was much greater than the variance within groups. This means that the variability due to differences of group means (the treatment effect *and* noise) was greater than the variability of individual scores from their own group mean (noise). So we can conclude that there is a treatment effect.

Stop Time: ______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Three

Using <u>purple pen</u>:

1) Correct your calculations as needed
2) <u>Copy</u> or <u>paraphrase</u> the explanations from Part Two onto Part Three

Re-writing correct explanations can improve your understanding of the material, and so it is very important you re-write them as close to verbatim as you can.

Turn in Parts Two & Three to your TA to get the Quiz

Start Time: ______________

One-Way ANOVA

The Dataset

	Group 1	Group 2	Group 3
	3	8	21
	5	6	16
	1	11	32
	0	7	16
	7	4	25
	2	12	40
$\bar{X}$	3	8	25

Fill in the Summary Table

Source	*df*	*SS*	*MS*	*F*
Across Group	**2**	**1596**	**798**	**22.48**
Within Group	**15**	**532**	**35.5**	
Total	**17**	**2128**		

Ho : $\mu_{Group1} = \mu_{Group2} = \mu_{Group3}$

SUMS OF SQUARES

Definitions	Calculations
x_{ij} = *individual score;* $\bar{x}_j$ = *group mean;* $\bar{x}_{..}$ = *grand mean* The Sum of Squares **total (SS_{TOT})** is the sum of the squared deviations of all the **individual scores** from the **grand (or overall) mean**. *This is a measure of the total amount of variability in the data.* $SS_{TOT} = SS_{AC} + SS_{WG}$	$SS_{TOT} = \Sigma(x_{ij} - \bar{x}_{..})^2$ (3-12)²+(5-12)²+(1-12)²+(0-12)²+(7-12)²+(2-12)²+(8-12)²+(6-12)²+(11-12)²+(7-12)²+(4-12)²+(12-12)²+(21-12)²+(16-12)²+(32-12)²+(16-12)²+(25-12)²+(40-12)² **= 2128**
The Sum of Squares **across groups (SS_{AG})** is the sum of the squared deviations of **each group mean** from the **grand (or overall) mean**, times the number of datapoints in that group. We multiply by the number of scores in the group to get a weighted estimate of the variability. *This is a measure of the amount of variability due to the treatment effect and due to noise.* SS_{AC} = *treatment effect + noise*	$SS_{AG} = (\Sigma(n(\bar{x}_j - \bar{x}_{..})^2))$ (6*(3-12)²)+(6*(8-12)²)+(6*(25-12)²) **= 1596**
The Sum of Squares **within groups (SS_{WG})** is the sum of the squared deviations of the **individual scores** from **their group mean**. *This is a measure of the amount of variability within each group due to noise.* SS_{WG} = *noise*	$SS_{WG} = \Sigma(\bar{x}_{ij} - \bar{x}_j)^2$ (3-3)²+(5-3)²+(1-3)²+(0-3)²+(7-3)²+(2-3)²+(8-8)²+(6-8)²+(11-8)²+(7-8)²+(4-8)²+(12-8)²+(21-25)²+(16-25)²+(32-25)²+(16-25)²+(25-25)²+(40-25)² **= 532**

Thought Questions

- **What main idea are these SSs capturing?**

- **Why would you expect SS_{AG} to be *different* when H_O is true rather than false? Consider whether you would expect SS_{AG} to be large or small when H_O is true and when H_O is false, explaining your reasoning.**

MEAN SQUARES & F-RATIO

Definitions	Calculations
k = Number of Groups; N = total number of datapoints The Mean Square **(MS)** is the amount of variance within each partition. MS has the same formula as variance (S^2): **MS** $= \frac{SS}{df} = \frac{SS}{n-1} = S^2$	$MS_{AG} = \frac{SS_{AG}}{df_{AG}} = \frac{1596}{2} =$ **798** df_{AG} = (k – 1) = (3 – 1) = **2**
It can be thought of as the *average* (or mean) variability per degree of freedom.	$MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{532}{15} =$ **35.5** df_{WG} = $df_{TOT} - df_{AC}$ = (N – 1) – (k – 1) = (18 – 1) – (3 – 1) = (17 – 2) = **15**
The F-ratio puts the amount of variability among the groups over the amount of variability within each group.	$F_{2,15} = \frac{MS_{AG}}{MS_{WG}} = \frac{798}{35.5} =$ **22.48**

Thought Questions

- **If H_O is true, is MS_{AG} smaller or larger than MS_{WG}? If H_O is false? Explain how you know.**

- **How does the F-ratio quantify the *treatment effect*?**

STATISTICAL CONCLUSION	
Definitions	**Calculations**
This is the critical value by which we judge if the amount of variability between groups is less than 5% likely to occur between groups from the same population.	$_{.05}F_{2,15} = 3.68$ *(from table)* Decision: **Reject** Fail to Reject
Thought Question	
• **Interpret your conclusion in terms of the treatment effect. Discuss variability *among* groups and *within* groups.**	

Did you finish?
____ Yes ____ No
Would you have wanted more time?
____ Yes ____ No

Stop Time: ______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Two

For each section:

1) Read through the definitions carefully, and fill in the blanks with the appropriate abbreviation
2) Read the calculations and fill in the blanks with the appropriate values (you do not have to calculate the answer once these are filled in)
3) Answer the thought questions to the best of your ability
4) Fill in the Summary Table as you solve each part

See your TA for Part Three

Start Time: _______________

One-Way ANOVA

The Dataset

	Group 1	Group 2	Group 3
	3	8	21
	5	6	16
	1	11	32
	0	7	16
	7	4	25
	2	12	40
$\bar{X}$	3	8	25

Fill in the Summary Table

Source	*df*	*SS*	*MS*	*F*
Across Group				
Within Group				
Total				

$$Ho : \mu_{Group1} = \mu_{Group2} = \mu_{Group3}$$

SUMS OF SQUARES

Definitions	Calculations
x_{ij} = *individual score;* $\bar{x}_j$ = *group mean;* $\bar{x}_{\_}$ = *grand mean* The Sum of Squares **total** (____) is the sum of the squared deviations of all the **individual scores** from the **grand (or overall) mean**. *This is a measure of the total amount of variability in the data.* $SS_{TOT} = SS_{AG} + SS_{WG}$	$SS_{TOT} = \Sigma(x_{ij} - \bar{x})^2$ $(3-\_\_\_)^2+(5-\_\_\_)^2+(1-\_\_\_)^2+(0-\_\_\_)^2+(7-\_\_\_)^2+(2-\_\_\_)^2+(8-\_\_\_)^2+(6-\_\_\_)^2+(11-\_\_\_)^2+(7-\_\_\_)^2+(4-\_\_\_)^2+(12-\_\_\_)^2+(21-\_\_\_)^2+(16-\_\_\_)^2+(32-\_\_\_)^2+(16-\_\_\_)^2+(25-\_\_\_)^2+(40-\_\_\_)^2$ **= 2128**
The Sum of Squares **across groups** (____) is the sum of the squared deviations of **each group mean** from the **grand (or overall) mean**, times the number of datapoints in that group. We multiply by the number of scores in the group to get a weighted estimate of the variability. *This is a measure of the amount of variability due to the treatment effect and due to noise.* SS_{AG} = *treatment effect + noise*	$SS_{AG} = (\Sigma(n(\bar{x}_j - \bar{x}_{\_})^2))$ $(\_\_\_*(\_\_\_-\_\_\_)^2)+(\_\_\_*(\_\_\_-\_\_\_)^2)+(\_\_\_*(\_\_\_-\_\_\_)^2)$ **= 1596**
The Sum of Squares **within groups** (____) is the sum of the squared deviations of the **individual scores** from **their group mean**. *This is a measure of the amount of variability within each group due to noise.* SS_{WG} = *noise*	$SS_{WG} = \Sigma(\bar{x}_{ij} - \bar{x}_j)^2$ $(3-\_\_\_)^2+(5-\_\_\_)^2+(1-\_\_\_)^2+(0-\_\_\_)^2+(7-\_\_\_)^2+(2-\_\_\_)^2+(8-\_\_\_)^2+(6-\_\_\_)^2+(11-\_\_\_)^2+(7-\_\_\_)^2+(4-\_\_\_)^2+(12-\_\_\_)^2+(21-\_\_\_)^2+(16-\_\_\_)^2+(32-\_\_\_)^2+(16-\_\_\_)^2+(25-\_\_\_)^2+(40-\_\_\_)^2$ **= 532**

Thought Questions

- **What main idea are these SSs capturing?**

- **Why would you expect SS_{AG} to be *different* when H_O is true rather than false? Consider whether you would expect SS_{AG} to be large or small when H_O is true and when H_O is false, explaining your reasoning.**

MEAN SQUARES & F-RATIO

Definitions	Calculations
k = Number of Groups; N = total number of datapoints The Mean Square **(MS)** is the amount of variance within each partition. MS has the same formula as variance (S^2): $MS = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom.	$MS_{AG} = \frac{SS_{AG}}{df_{AG}} = \frac{\_\_\_}{\_\_\_} =$ **798** $df_{AG} = (k-1) = (\_\_\_-1) =$ **2** $MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{\_\_\_}{\_\_\_} =$ **35.5** $df_{WG} = df_{TOT} - df_{AG} = (N-1)-(k-1)$ $= (\_\_\_-1)-(\_\_\_-1) = (\_\_\_-\_\_\_) =$ **15**
The F-ratio puts the amount of variability among the groups over the amount of variability within each group.	$F_{2,15} = \frac{MS_{AG}}{MS_{WG}} = \frac{\_\_\_}{\_\_\_} =$ **22.48**

Thought Questions

- **If H_O is true, is MS_{AG} smaller or larger than MS_{WG}? If H_O is false? Explain how you know.**

- **How does the F-ratio quantify the *treatment effect*?**

STATISTICAL CONCLUSION	
Definitions	**Calculations**
This is the critical value by which we judge if the amount of variability between groups is less than 5% likely to occur between groups from the same population.	${}_{.05}F_{2,15} = 3.68$ *(from table)* Decision: **Reject** Fail to Reject
Thought Question	
• **Interpret your conclusion in terms of the treatment effect. Discuss variability *among* groups and *within* groups.**	

Stop Time: ______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Three

Using purple pen, use Part 3 to do back and correct your work in Part 2:

1) Correct the blanks you filled in as needed, including the Summary Table
2) Edit or improve your answers to the thought questions using the answers that follow. Do not just copy the explanations over; it is important you modify or edit what you already have written. Be sure you include all the main ideas in the provided thought question answers.

Turn in Parts Two & Three to your TA to get the Quiz

Start Time: ______________

One-Way ANOVA

The Dataset

	Group 1	Group 2	Group 3
	3	8	21
	5	6	16
	1	11	32
	0	7	16
	7	4	25
	2	12	40
$\bar{X}$	3	8	25

Fill in the Summary Table

Source	*df*	*SS*	*MS*	*F*
Across Group	**2**	**1596**	**798**	**22.48**
Within Group	**15**	**532**	**35.5**	
Total	**17**	**2128**		

Ho : $\mu_{Group1} = \mu_{Group2} = \mu_{Group3}$

SUMS OF SQUARES

Definitions	Calculations
x_{ij} = *individual score;* $\bar{x}_j$ = *group mean;* $\bar{x}_{\_}$ = *grand mean* The Sum of Squares **total (SS_{TOT})** is the sum of the squared deviations of all the **individual scores** from the **grand (or overall) mean**. *This is a measure of the total amount of variability in the data.* $SS_{TOT} = SS_{AC} + SS_{WG}$	$SS_{TOT} = \Sigma(x_{ij} - \bar{x})^2$ $(3-12)^2+(5-12)^2+(1-12)^2+(0-12)^2+(7-12)^2+(2-12)^2+(8-12)^2+(6-12)^2+(11-12)^2+(7-12)^2+(4-12)^2+(12-12)^2+(21-12)^2+(16-12)^2+(32-12)^2+(16-12)^2+(25-12)^2+(40-12)^2$ **= 2128**
The Sum of Squares **across groups (SS_{AG})** is the sum of the squared deviations of **each group mean** from the **grand (or overall) mean**, times the number of datapoints in that group. We multiply by the number of scores in the group to get a weighted estimate of the variability. *This is a measure of the amount of variability due to the treatment effect and due to noise.* SS_{AC} = *treatment effect + noise*	$SS_{AG} = (\Sigma(n(\bar{x}_j - \bar{x})^2))$ $(6*(3-12)^2)+(6*(8-12)^2)+(6*(25-12)^2)$ **= 1596**
The Sum of Squares **within groups (SS_{WG})** is the sum of the squared deviations of the **individual scores** from **their group mean**. *This is a measure of the amount of variability within each group due to noise.* SS_{WG} = *noise*	$SS_{WG} = \Sigma(\bar{x}_{ij} - \bar{x}_j)^2$ $(3-3)^2+(5-3)^2+(1-3)^2+(0-3)^2+(7-3)^2+(2-3)^2+(8-8)^2+(6-8)^2+(11-8)^2+(7-8)^2+(4-8)^2+(12-8)^2+(21-25)^2+(16-25)^2+(32-25)^2+(16-25)^2+(25-25)^2+(40-25)^2$ **= 532**

Thought Questions

- **What main idea are these SSs capturing?**

The main idea is that we are capturing the *amount of variability* in the data. SS_{TOT} captures all of it, and SS_{AG} and SS_{WG} partition this variability into parts that are due to the treatment effect and noise.

- **Why would you expect SS_{AG} to be *different* when H_O is true rather than false? Consider whether you would expect SS_{AG} to be large or small when H_O is true and when H_O is false, explaining your reasoning.**

The value of SS_{AG} would be different if H_O were true or false, because SS_{AG} captures how different each group mean is from the grand mean. If H_O were true, the group means would be very similar to the grand mean, resulting in small differences, and a small value for the SS_{AG}. If H_O were false, the group means would be different from the grand mean, resulting in larger differences, and a larger value for the SS_{AG}.

MEAN SQUARES & F-RATIO

Definitions	Calculations
k = Number of Groups; N = total number of datapoints The Mean Square **(MS)** is the amount of variance within each partition. MS has the same formula as variance (S^2): $MS = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom.	$MS_{AG} = \frac{SS_{AG}}{df_{AG}} = \frac{1596}{2} = $ **798** df_{AG} = (k – 1) = (3 – 1) = **2** $MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{532}{15} = $ **35.5** df_{WG} = $df_{TOT} - df_{AG}$ = (N – 1) – (k – 1) = (18 – 1) – (3 – 1) = (17 – 2) = **15**
The F-ratio puts the amount of variability among the groups over the amount of variability within each group.	$F_{2,15} = \frac{MS_{AG}}{MS_{WG}} = \frac{798}{35.5} = $ **22.48**

Thought Questions

- **If H_O is true, is MS_{AG} smaller or larger than MS_{WG}? If H_O is false? Explain how you know.**

If H_O is true, MS_{AG} and MS_{WG} should be approximately equal. MS_{AG} includes variability due to the treatment effect and due to noise. Since there is no treatment effect, MS_{AG} would only include variability due to noise. MS_{WG} only includes variability due to noise anyway, so they should be about the same.
If H_O is false, MS_{AG} should be larger than the MS_{WG}. MS_{AG} includes variability due to the treatment effect *and* due to noise. Since there is a treatment effect, MS_{AG} would include this variability *plus* the variability due to noise. MS_{WG} only includes variability due to noise. So, MS_{AG} should be larger.

- **How does the F-ratio quantify the *treatment effect*?**

The F-ratio divides the MS_{AG} by the MS_{WG}. The MS_{AG} contains variability due to the treatment effect *and* noise, and dividing it by the MS_{WG}, which only contains variably due to noise, will cancel out the variability due to noise. The resulting value of the F-ratio reflects the size of variability due to the treatment effect.

STATISTICAL CONCLUSION

Definitions	Calculations
This is the critical value by which we judge if the amount of variability between groups is less than 5% likely to occur between groups from the same population.	${}_{.05}F_{2,15}$ = 3.68 *(from table)* Decision: **Reject** Fail to Reject

Thought Question

- **Interpret your conclusion in terms of the treatment effect. Discuss variability *among* groups and *within* groups.**

There is a treatment effect, and not all of the group means are the same. The variance among groups was much greater than the variance within groups. This means that the variability due to differences of group means (the treatment effect *and* noise) was greater than the variability of individual scores from their own group mean (noise). So we can conclude that there is a treatment effect.

Did you finish?
____ Yes ____ No
Would you have wanted more time?
____ Yes ____ No

Stop Time: ______________

Name: ______________________________

Quiz Time: **20 minutes**

Rating Items – 3 minutes

Quiz Items – 17 minutes or until end of class

Start Time: __________

Section Time: **3 minutes**

Think about the calculations and thought questions you just completed. Please answer the following questions about your work with them in mind.

Please answer the following items on a scale from 1(low) to 5 (high), with 3 being the neutral point.

1)	In solving or studying these worksheets I invested _____ mental effort	Low	1	2	3	4	5	High
2)	How easy or difficult was it to complete these worksheets?	Easy	1	2	3	4	5	Difficult
3)	How mentally demanding were these worksheets?	Low	1	2	3	4	5	High
4)	How hurried or rushed was the pace of these worksheets?	Not Rushed	1	2	3	4	5	Very Rushed
5)	How successful were you in accomplishing what you were asked to do?	Not At All	1	2	3	4	5	Very
6)	How hard did you have to work to accomplish your level of performance?	Not At All	1	2	3	4	5	Very
7)	How insecure, discouraged, irritated, stressed, and annoyed were you?	Not At All	1	2	3	4	5	Very
8)	The calculations helped me to understand ANOVA.	Not At All	1	2	3	4	5	Very
9)	I will think of these calculations when solving ANOVA problems in the future	Not At All	1	2	3	4	5	Very
10)	The definitions helped me to understand ANOVA	Not At All	1	2	3	4	5	Very
11)	I will think of these definitions when solving ANOVA problems in the future	Not At All	1	2	3	4	5	Very
12)	The thought-questions helped me to understand ANOVA	Not At All	1	2	3	4	5	Very
13)	I will think of these thought-questions when solving ANOVA problems in the future	Not At All	1	2	3	4	5	Very

For the following items, response to if you engaged in these activities either <u>in your head</u> or <u>on paper</u>, whether you were instructed to or not.

Use the following scale:

1- Rarely while working 2- Somewhat often while working 3- About half the time while working 4- Quite often while working 5- Almost always while working

		Rarely				Almost Always
14)	I thought deeply about what my calculated values represented and meant	1	2	3	4	5
15)	I identified the most important ideas	1	2	3	4	5
16)	I copied the thought-questions exactly	1	2	3	4	5
17)	I paraphrased the thought-questions	1	2	3	4	5
18)	I focused my attention on the thought-questions	1	2	3	4	5
19)	I explained in my own words what the thought-questions meant to me	1	2	3	4	5
20)	I developed my own answers to the thought-questions before reading the provided answers	1	2	3	4	5
21)	My own answers to the thought-questions included the ideas in the provided answers	1	2	3	4	5
22)	I connected the thought-questions to ideas I already knew	1	2	3	4	5
23)	I made hypotheses or predictions about the thought-questions	1	2	3	4	5
24)	I justified or provided reasons why concepts in the thought-questions occur	1	2	3	4	5

Stop Time: ______________

Start Time: ______________

Section Time: **17 minutes**

1) The F-value of an ANOVA is 2.92, and the MS_{WG} is 7.32.

Find the MS_{AG}: ______________

Without knowing anything else, what could you suppose is the case about H_O? True? Not true?

Why do you think so? (be specific):

2) What happens to the value of the F-ratio if the differences between the sample means increases? Why?

3) Complete the summary table:

k (# groups) = 4
n per group = 10

Source	Df	SS	MS	F
Across Group			353.267	
Within Group		238.6		
Total	39			

What's your decision? (circle one): Reject H_O Fail to Reject H_O

4) Circle the answer that best fills in the blanks:

The ___________ the value of the F-ratio, the ______________ likely it is that the sample means from the groups represent one population mean.

a. smaller; less b. larger; less c. larger; more

5) Explain why the expected value for an F-ratio is equal to about 1 when there is no treatment effect?

6) Look at the data below. Without doing any calculations on paper, try to predict what values should be obtained for SS_{AG}, MS_{AG}, and the F-ratio.

Group 1	Group 2
4	2
0	1
1	0
3	5

Predicted Values:

SS_{AG}: __________ MS_{AG}: __________ F-ratio : __________

7) Consider a dataset with five samples that are all the same size, but whose standard deviations are different:

$sd_1 = 10$ $sd_2 = 15$ $sd_3 = 12$ $sd_4 = 11$ $sd_5 = 10$

Describe in words how you would you calculate the value of MS_{WG} with this information.

Stop Time: ______________

Did you finish?
____ Yes ____ No
Would you have wanted more time?
____ Yes ____ No

Name: ____

1) When the amount of variability within groups increases, but the variability among groups stay the same, the value of the F-ratio ______________________________.

Why?

2) Why would you expect SS_{AG} to be *different* when H_O is true rather than false? Consider whether you would expect SS_{AG} to be large or small when H_O is true and when H_O is false, explaining your reasoning.

3) Complete the summary table:

k (# groups) = 3
n per group = 10

Source	Df	SS	MS	F
Across Group		1660		83.0
Within Group			10	
Total	29			

Without knowing anything else, what could you suppose the decision would be?
(circle one): Reject H_O Fail to Reject H_O

Why do you think so? (be specific)

4)

	Treats Earned Per Week	
	Cats	Dogs
	1	10
	2	20
	3	30
$\bar{X}$	**2**	**20**
$\bar{X}_{..}$	**11**	

Listed above is some data on how many treats per week cats and dogs earn from their owners by doing tricks. Showing your work, perform an ANOVA by determining the following:

SS_{TOT} : ** Do not have to calculate final value for this, just show how you would get there.*	
SS_{AG} :	
SS_{WG} :	
MS_{AG} :	
MS_{WG} :	
F-Ratio :	
$_{.05}F_{\_\_,\_\_}$:	**7.71**
Decision:	Reject Fail To Reject

Name: ____

Section Time: **3 minutes**

Keep your answers brief

Wait until your TA says to begin

Start Time: _______________

Section Time: 3 minutes

1) At a high level, how does a Two-Way ANOVA determine if there are differences between groups?

2) Briefly list the steps needed calculate a Two-Way ANOVA.

3) How do the steps involved in a Two-Way ANOVA answer the question we are trying to test?

Stop Time: _______________

Name: ______________________________

Section Time: **25 minutes for Parts Two & Three**

Part Two

For each section:

5) In column 1, read through the definitions carefully, and fill in the blanks with the appropriate abbreviation
6) In column 2, complete the calculations
7) Read the thought questions carefully
8) Fill in the Summary Table once you have completed the worksheet

See your TA for Part Three

Start Time: __________

Two-Way ANOVA

In a study of group dynamics, 30 people were grouped into teams of 5, and were given directions for a group project that had them work cooperatively, neutrally, or competitively. There were 3 teams of women and 3 teams of men, each in one of the group dynamic conditions. This study has two factors: **Gender** (2 levels: women and men) and **Group Dynamic** (3 levels: cooperative, neutral, and competitive). They were then asked to rate their enjoyment of the project on a scale from 0-100, with 100 being highly enjoyable. Is there a difference in **task enjoyment** between **genders**? Between the different **dynamic** groups? Is there a **gender by dynamic interaction**?

THE DATASET

	Cooperative	Neutral	Competitive
Women	**95**	**75**	**35**
$\bar{X}_{Women}$ =	98	79	38
68.3	96	71	32
	84	66	26
	100	84	44
	97	75	35
Men	**40**	**50**	**80**
$\bar{X}_{Men}$ =	44	41	75
56.7	31	55	89
	36	50	80
	49	59	85
	40	45	71
$\bar{X}_{Total}$ = **62.5**	$\bar{X}_{Coop}$ = **67.5**	$\bar{X}_{Neut}$ = **62.5**	$\bar{X}_{Comp}$ = **57.5**

ΣX / T VALUES

	Cooperative	Neutral	Competitive	
Women	475	375	175	**1025**
Men	200	250	400 $_{Cells}$	**850** $_{Rows}$
	675	**625**	**575** $_{Columns}$	**1875** $_{Total}$
	ΣX^2 = **133027**			

NULL HYPOTHESES

H_O Gender:	$\mu_{Women} = \mu_{Men}$
H_O Dynamic:	$\mu_{Coop} = \mu_{Neut} = \mu_{Comp}$
H_O Interaction:	*No Interaction*

SUMS OF SQUARES

Definitions	Calculations
$T_{group} = \Sigma X_{group}$ **n** = # in Each Group **c** = # Columns **r** = # Rows The sum of squares **total** (_____) is a measure of the total amount of variability in the data. $SS_{TOT} = SS_{AG} + SS_{WG}$	$SS_{TOT} = \Sigma X^2 - \frac{T^2_{Total}}{N_{Total}}$ $133027 - \frac{1875^2}{30}$ = 133027 – ____________ = ____________
The sum of squares **within groups** (_____) is a measure of the amount of variability within each group due to noise. SS_{WG} = *noise*	$SS_{WG} = \Sigma X^2 - \Sigma(\frac{T^2_{cell}}{n_{cell}})$ $133027 - \frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5}$ = 133027 – ____________ = ____________
The sum of squares **among** each of the 6 **groups** or cells (______) is a measure of the amount of variability due to the treatment effect and due to noise. Also called the SS_{Cells} SS_{AG} = ____________ *effect + noise* The SS_{AG} can be further partitioned into the SS of row, column, and interaction.	$SS_{AG} = \Sigma(\frac{T^2_{cell}}{n_{cell}}) - \frac{T^2_{Total}}{N_{Total}}$ $\frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5} - \frac{1875^2}{30}$ = ____________ – ____________

$SS_{AG} = SS_{Gender} + SS_{Dynamic} + SS_{Interaction}$	= ____________
The Sum of Squares of **Gender** (______) is the amount of variability across each of the 2 gender rows. SS_{Gender} = *effect of* ____________ *+ noise*	$SS_{Gender} = \Sigma(\frac{T^2_{Gender}}{n_{Gender}}) - \frac{T^2_{Total}}{N_{Total}}$ $= \frac{1025^2}{15} + \frac{850^2}{15} - \frac{1875^2}{30}$ = ____________ – ____________ = ____________
The Sum of Squares of Group **Dynamic** (________) is the amount of variability across each of the 3 dynamic columns. $SS_{Dynamic}$ = *effect of* ____________ *+ noise*	$SS_{Dynamic} = \Sigma(\frac{T^2_{Dynamic}}{n_{Dynamic}}) - \frac{T^2_{Total}}{N_{Total}} = \frac{675^2}{10} + \frac{625^2}{10} + \frac{575^2}{10} - \frac{1875^2}{30}$ = ____________ – ____________ = ____________
The sum of squares of the **interaction** (____) is remaining variability among the groups that is not due to the effect of row and of column. Therefore, it must be due to the effect of the interaction between the row and column. $SS_{Interaction}$ = *effect of* ____________ *+ noise*	$SS_{Interaction} = SS_{AG} - SS_{Gender} - SS_{Dynamic}$ = ____________ – ____________ – ____________ = ____________

Thought Questions

- **The total variability is partitioned into each of these Sums of Squares (SSs). Describe how each SS listed above 'fits' together to form the SS_{TOTAL}.**

  The SS total is made up of the SS within group and among group. The SS_{AG} is made up of the SS of the row, the column, and the interaction.

$$SS_{TOTAL} = SS_{WG} + SS_{AG}$$
$$SS_{ROW} + SS_{COL} + SS_{INTERACT}$$

MEAN SQUARES & F-RATIO

Definitions	Calculations
k *= # Groups;* ***N*** *= Total # Datapoints* **c** = # Columns **r** = # Rows The Mean Square (___) is the amount of variance within each partition. MS has the same formula as variance (S^2): $MS = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom.	$MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{1152}{24}$ = ____________ $df_{WG} = df_{TOT} - df_{AG}$ = (N – 1) – (k – 1) = (30 – 1) – (6 – 1) = (29 – 5) = ____________
	$MS_{Gender} = \frac{SS_{Gender}}{df_{Gender}} = \frac{1020.83}{1}$ = ____________ df_{Gender} = (r – 1) = (2 – 1) = ____________
	$MS_{Dynamic} = \frac{SS_{Dynamic}}{df_{Dynamic}} = \frac{500}{2}$ = ____________ $df_{Dynamic}$ = (c – 1) = (3 – 1) = ____________

The **F-ratio** puts the amount of variability among the rows, columns, or the interaction over the amount of variability *within* each group.	$MS_{Inter} = \frac{SS_{Inter}}{df_{Inter}} = \frac{13166.67}{2}$ = ____________ $df_{Inter} = df_{Gender} * df_{Dynamic}$ = (1 * 2) = ____________
The Gender Effect: Differential effects produced by different levels of the row factor.	$F_{Gender\ dfGender,dfWG} = F_{Gender\_\_,\_\_}$ $= \frac{MS_{Gender}}{MS_{WG}} = \frac{1020.83}{48}$ = ____________
The Dynamic Effect: Differential effects produced by different levels of the column factor.	$F_{Dynamic\ dfDynamic,dfWG} = F_{Dynamic\_\_,\_\_}$ $= \frac{MS_{Dynamic}}{MS_{WG}} = \frac{250}{48}$ = ____________
The Interaction Effect: Differences that are produced by unique combinations of gender and dynamic. An interaction is present when the effects of the row factors change with different levels of the column factor.	$F_{Inter\ dfInter,dfWG} = F_{Inter\_\_,\_\_}$ $= \frac{MS_{Inter}}{MS_{WG}} = \frac{626.98}{48}$ = ____________

Thought Questions

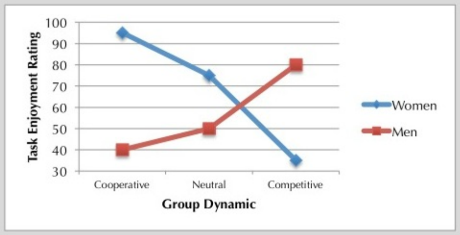
- **If H_O of gender is false (there *are* differences between genders), would you expect MS_{Gender} to be the same or larger than MS_{WG}? Explain how you know.**

  If H_O about gender is false, MS_{Gender} should be larger than MS_{WG}. MS_{Gender} includes variability due to the effect of gender *and* noise. Since there is a gender effect, MS_{Gender} would include this *plus* the variability due to noise, whereas MS_{WG} only includes variability due to noise.

STATISTICAL CONCLUSION

Definitions	Calculations
The Gender Effect: This F-ratio tests the null hypothesis that there is no main effect of ____________.	$F_{Gender\_\_,\_\_}$ = ____________ $_{.05}F_{crit\ \_\_,\_\_}$ = 4.26 Gender Main Effect Decision: **Reject** Fail to Reject
The Dynamic Effect: This F-ratio tests the null hypothesis that there is no main effect of ____________.	$F_{Dynamic\_\_,\_\_}$ = ____________ $_{.05}F_{crit\ \_\_,\_\_}$ = 3.40 Dynamic Main Effect Decision: **Reject** Fail to Reject
The Interaction Effect: This F-ratio tests the null hypothesis that there is no gender by group dynamic ____________ effect.	$F_{Inter\_\_,\_\_}$ = ____________ $_{.05}F_{crit\ \_\_,\_\_}$ = 3.40 Interaction Effect Decision: **Reject** Fail to Reject

Graph of the means of each group:



Thought Question

- **What can you conclude about the task enjoyment of the people? Interpret your conclusion in terms of the main effects of gender, of dynamic, and of the interaction effect. The graph may help.**

There is an effect of **gender**, meaning there are differences in enjoyment between gender, regardless of dynamic. There is also an effect of **group dynamic**, meaning there are differences in enjoyment between dynamics, regardless of gender. There is also an **interaction** between gender and dynamic on enjoyment: the effect of dynamic depends on gender. Women preferred the cooperative dynamic most and disfavored the competitive; whereas the men preferred the competitive dynamic most and disfavored the cooperative.

- **Why could it be problematic to interpret the main effects of gender and dynamic (even if they were significant) when the interaction is significant?**

Men and women (regardless of dynamic) do differ in task enjoyment, and the three dynamic groups (regardless of gender) differ in task enjoyment as well. However, since there is an interaction, the effect is **not consistent**. This makes it difficult to say something conclusive about the individual main effects. For example, we cannot conclude that being cooperative was always preferred, because it was for women, but not for men.

SUMMARY TABLE

Source of Variation	df	SS	MS	F	$_{.05}F_{crit}$	Decision
Within Groups	**24**					
Among Groups	**(5)**					
Gender					4.26	
Dynamic					3.40	
Gender x Dynamic					3.40	
TOTAL	**29**					

Stop Time: ______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Three

Using purple pen:

3) Correct your calculations as needed

4) Copy or paraphrase the explanations from Part Two onto Part Three

Re-writing correct explanations can improve your understanding of the material, and so it is very important you re-write them as close to verbatim as you can.

Turn in Parts Two & Three to your TA to get the Quiz

Start Time: _______________

Two-Way ANOVA

In a study of group dynamics, 30 people were grouped into teams of 5, and were given directions for a group project that had them work cooperatively, neutrally, or competitively. There were 3 teams of women and 3 teams of men, each in one of the group dynamic conditions. This study has two factors: **Gender** (2 levels: women and men) and **Group Dynamic** (3 levels: cooperative, neutral, and competitive). They were then asked to rate their enjoyment of the project on a scale from 0-100, with 100 being highly enjoyable. Is there a difference in **task enjoyment** between **genders**? Between the different **dynamic** groups? Is there a **gender by dynamic interaction**?

THE DATASET

	Cooperative	Neutral	Competitive
Women	**95**	**75**	**35**
$\bar{X}_{Women}$ =	98	79	38
68.3	96	71	32
	84	66	26
	100	84	44
	97	75	35
Men	**40**	**50**	**80**
$\bar{X}_{Men}$ =	44	41	75
56.7	31	55	89
	36	50	80
	49	59	85
	40	45	71
$\bar{X}_{Total}$ = **62.5**	$\bar{X}_{Coop}$ = **67.5**	$\bar{X}_{Neut}$ = **62.5**	$\bar{X}_{Comp}$ = **57.5**

ΣX / T VALUES

	Cooperative	Neutral	Competitive	
Women	475	375	175	**1025**
Men	200	250	400 $_{Cells}$	**850** $_{Rows}$
	675	**625**	**575** $_{Columns}$	**1875** $_{Total}$
ΣX^2 = **133027**				

NULL HYPOTHESES

H_O Gender:	$\mu_{Women} = \mu_{Men}$
H_O Dynamic:	$\mu_{Coop} = \mu_{Neut} = \mu_{Comp}$
H_O Interaction:	*No Interaction*

SUMS OF SQUARES

Definitions	Shortcut Formula
$\mathbf{T_{group}} = \Sigma X_{group}$ **n** = # in Each Group **c** = # Columns **r** = # Rows The sum of squares **total** (SS_{TOT}) is a measure of the total amount of variability in the data. $SS_{TOT} = SS_{AC} + SS_{WG}$	$\mathbf{SS_{TOT}} = \Sigma X^2 - \frac{T^2_{Total}}{N_{Total}}$ $133027 - \frac{1875^2}{30}$ $= 133027 - 117187.5$ **= 15839.5**
The sum of squares **within groups** (SS_{WG}) is a measure of the amount of variability within each group due to noise. SS_{WG} = *noise*	$\mathbf{SS_{WG}} = \Sigma X^2 - \Sigma(\frac{T^2_{cell}}{n_{cell}})$ $133027 - \frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5}$ $= 133027 - 131875$ **= 1152**
The sum of squares **among** each of the 6 **groups** or cells (SS_{AG}) is a measure of the amount of variability due to the treatment effect and due to noise. $SS_{AG} = SS_{Cells}$ SS_{AG} = _______________ *effect + noise* The SS_{AG} can be further partitioned into the SS of row, column, and interaction. $SS_{AG} = SS_{Gender} + SS_{Dynamic} + SS_{Interaction}$	$\mathbf{SS_{AG}} = \Sigma(\frac{T^2_{cell}}{n_{cell}}) - \frac{T^2_{Total}}{N_{Total}}$ $\frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5} - \frac{1875^2}{30}$ $= 131875 - 117187.5$ **= 14687.5**

The Sum of Squares of **Gender** ($\mathbf{SS_{Gender}}$) is the amount of variability across each of the 2 gender rows. SS_{Gender} = *effect of* _______________ *+ noise*	$\mathbf{SS_{Gender}} = \Sigma(\frac{T^2_{Gender}}{n_{Gender}}) - \frac{T^2_{Total}}{N_{Total}}$ $= \frac{1025^2}{15} + \frac{850^2}{15} - \frac{1875^2}{30}$ $= 118208.3 - 117187.5$ **= 1020.83**
The Sum of Squares of Group **Dynamic** ($\mathbf{SS_{Dynamic}}$) is the amount of variability across each of the 3 dynamic columns. $SS_{Dynamic}$ = *effect of* _______________ *+ noise*	$\mathbf{SS_{Dynamic}} = \Sigma(\frac{T^2_{Dynamic}}{n_{Dynamic}}) - \frac{T^2_{Total}}{N_{Total}} = \frac{675^2}{10} + \frac{625^2}{10} + \frac{575^2}{10} - \frac{1875^2}{30}$ $= 117687.5 - 117187.5$ **= 500**
The sum of squares of the **interaction** ($\mathbf{SS_{Interaction}}$) is remaining variability among the groups that is not due to the effect of row and of column. Therefore, it must be due to the effect of the interaction between the row and column. $\mathbf{SS_{Interaction}}$ = *effect of* _______________ *+ noise*	$\mathbf{SS_{Interaction}} = SS_{AG} - SS_{Gender} - SS_{Dynamic}$ $= 14687.5 - 1020.83 - 500$ **= 13166.67**

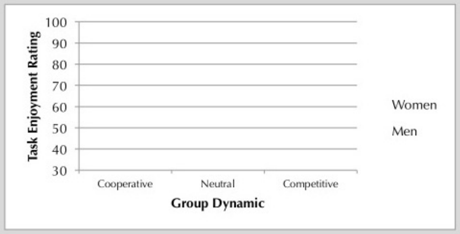
Thought Questions

- **The total variability is partitioned into each of these Sums of Squares (SSs). Describe how each SS listed above 'fits' together to form the SS_{TOTAL}.**

MEAN SQUARES & F-RATIO

Definitions	Calculations
k = *# Groups;* ***N*** = *Total # Datapoints* **c** = # Columns **r** = # Rows The Mean Square **(MS)** is the amount of variance within each partition. MS has the same formula as variance (S^2): $\mathbf{MS} = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom.	$\mathbf{MS_{WG}} = \frac{SS_{WG}}{df_{WG}} = \frac{1152}{24} =$ **48** $\mathbf{df_{WG}} = df_{TOT} - df_{AG} = (N-1) - (k-1)$ $= (30-1) - (6-1) = (29-5) =$ **24**
	$\mathbf{MS_{Gender}} = \frac{SS_{Gender}}{df_{Gender}} = \frac{1020.83}{1} =$ **1020.83** $\mathbf{df_{Gender}} = (r-1) = (2-1) =$ **1**
	$\mathbf{MS_{Dynamic}} = \frac{SS_{Dynamic}}{df_{Dynamic}} = \frac{500}{2} =$ **250** $\mathbf{df_{Dynamic}} = (c-1) = (3-1) =$ **2**

Graph of the means of each group:



Thought Question

- **What can you conclude about the task enjoyment of the people? Interpret your conclusion in terms of the main effects of gender, of dynamic, and of the interaction effect. The graph may help.**

- **Why could it be problematic to interpret the main effects of gender and dynamic (even if they were significant) when the interaction is significant?**

SUMMARY TABLE						
Source of Variation	**df**	**SS**	**MS**	**F**	$_{.05}F_{crit}$	**Decision**
Within Groups	**24**	1152	48			
Among Groups	**(5)**	(14687.5)				
Gender	1	1020.83	1020.83	21.267	$_{.05}F_{1,24} = 4.26$	Reject
Dynamic	2	500	250	5.208	$_{.05}F_{2,24} = 3.40$	Reject
Gender x Dynamic	2	13166.67	626.98	13.062	$_{.05}F_{2,24} = 3.40$	Reject
TOTAL	**29**	15839.5				

Stop Time: ____________

Did you finish?
____ Yes ____ No
Would you have wanted more time?
____ Yes ____ No

Name: ______________________________

Section Time: **25 minutes for Parts Two & Three**

Part Two

For each section:

5) Read through the definitions carefully, and fill in the blanks with the appropriate abbreviation

6) Read the calculations and fill in the blanks with the appropriate values (you do not have to calculate the answer once these are filled in)

7) Answer the thought questions to the best of your ability

8) Fill in the Summary Table once you have completed the worksheet.

Attempting to answer thought questions first can improve your memory of them, and so it is very important you do your best to answer the thought questions well and to the best of your ability in Part Two.

See your TA for Part Three

Start Time: ____________

Two-Way ANOVA

In a study of group dynamics, 30 people were grouped into teams of 5, and were given directions for a group project that had them work cooperatively, neutrally, or competitively. There were 3 teams of women and 3 teams of men, each in one of the group dynamic conditions. This study has two factors: **Gender** (2 levels: women and men) and **Group Dynamic** (3 levels: cooperative, neutral, and competitive). They were then asked to rate their enjoyment of the project on a scale from 0-100, with 100 being highly enjoyable. Is there a difference in **task enjoyment** between **genders**? Between the different **dynamic** groups? Is there a **gender by dynamic interaction**?

THE DATASET

	Cooperative	Neutral	Competitive
Women	**95**	**75**	**35**
$\bar{X}_{Women}$ =	98	79	38
68.3	96	71	32
	84	66	26
	100	84	44
	97	75	35
Men	**40**	**50**	**80**
$\bar{X}_{Men}$ =	44	41	75
56.7	31	55	89
	36	50	80
	49	59	85
	40	45	71
$\bar{X}_{Total}$ = **62.5**	$\bar{X}_{Coop}$ = **67.5**	$\bar{X}_{Neut}$ = **62.5**	$\bar{X}_{Comp}$ = **57.5**

ΣX / T VALUES

	Cooperative	Neutral	Competitive	
Women	475	375	175	**1025**
Men	200	250	400 $_{Cells}$	**850** $_{Rows}$
	675	**625**	**575** $_{Columns}$	**1875** $_{Total}$
ΣX^2 = **133027**				

NULL HYPOTHESES

H_O Gender:	$\mu_{Women} = \mu_{Men}$
H_O Dynamic:	$\mu_{Coop} = \mu_{Neut} = \mu_{Comp}$
H_O Interaction:	*No Interaction*

SUMS OF SQUARES

Definitions	Shortcut Formula
$T_{group} = \Sigma X_{group}$ **n** = # in Each Group **c** = # Columns **r** = # Rows The sum of squares **total** (____) is a measure of the total amount of variability in the data. $SS_{TOT} = SS_{AG} + SS_{WG}$	$SS_{TOT} = \Sigma X^2 - \frac{T^2_{Total}}{N_{Total}}$ ______ $-$ $\frac{\_\_\_\_^2}{\_\_\_\_}$ = ______ – 117187.5 = **15839.5**
The sum of squares **within groups** (____) is a measure of the amount of variability within each group due to noise. $SS_{WG} = noise$ The sum of squares **among** each of the 6 **groups** or cells (____) is a measure of the amount of variability due to the treatment effect and due to noise. Also called the SS of cells SS_{AG} = ______ *effect + noise*	$SS_{WG} = \Sigma X^2 - \Sigma(\frac{T^2_{cell}}{n_{cell}})$ ______ $-$ $\frac{\_\_\_^2}{\_\_\_} + \frac{\_\_\_^2}{\_\_\_} + \frac{\_\_\_^2}{\_\_\_} + \frac{\_\_\_^2}{\_\_\_} +$ $\frac{\_\_\_^2}{\_\_\_} + \frac{\_\_\_^2}{\_\_\_}$ = ______ – 131875 = **1152**

The SS_{AG} can be further partitioned into the SS of row, column, and interaction. $SS_{AG} = SS_{Gender} + SS_{Dynamic} + SS_{Interaction}$ The Sum of Squares of **Gender** (________) is the amount of variability across each of the 2 gender rows. SS_{Gender} = *effect of* ______________ + *noise* The Sum of Squares of Group **Dynamic** (_________) is the amount of variability across each of the 3 dynamic columns. $SS_{Dynamic}$ = *effect of* ______________ + *noise*	$SS_{AG} = \Sigma\left(\frac{T_{cell}^2}{n_{cell}}\right) - \frac{T_{Total}^2}{N_{Total}}$ $\frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} +$ $\frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} - \frac{\_\_\_\_^2}{\_\_\_\_}$ = 131875 – 117187.5 **= 14687.5** $SS_{Gender} = \Sigma\left(\frac{T_{Gender}^2}{n_{Gender}}\right) - \frac{T_{Total}^2}{N_{Total}}$ $= \frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} - \frac{\_\_\_\_^2}{\_\_\_\_}$ = 118208.3 - 117187.5 **= 1020.83** $SS_{Dynamic} = \Sigma\left(\frac{T_{Dynamic}^2}{n_{Dynamic}}\right) - \frac{T_{Total}^2}{N_{Total}}$ $= \frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} + \frac{\_\_\_\_^2}{\_\_\_\_} - \frac{\_\_\_\_^2}{\_\_\_\_}$ = 117687.5 - 117187.5 **= 500**
The sum of squares of the **interaction** (_________) is remaining variability among the groups that is not due to the effect of row and of column. Therefore, it must be due to the effect of the interaction between the row and column. $SS_{Interaction}$ = *effect of* ______________ + *noise*	$SS_{Interaction} = SS_{AG} - SS_{Gender} - SS_{Dynamic}$ = ____________ – ____________ – ____________ **= 13166.67**

Thought Questions

- **The total variability is partitioned into each of these Sums of Squares (SSs). Describe how each SS listed above 'fits' together to form the SS_{TOTAL}.**

MEAN SQUARES & F-RATIO

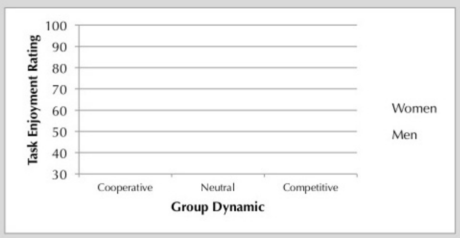
Definitions	Calculations
k = *# Groups*; ***N*** = *Total # Datapoints* **c** = # Columns **r** = # Rows The Mean Square (___) is the amount of variance within each partition. MS has the same formula as variance (S^2): $MS = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom. The **F-ratio** puts the amount of variability among the rows, columns, or the interaction over the amount of variability within each group. **The Gender Effect:** Differential effects produced by different levels of the row factor. **The Dynamic Effect:** Differential effects produced by different levels of the column factor. **The Interaction Effect:** Differences that are produced by unique combinations of ________________ and ________________. An interaction is present when the effects of the row factors change with different levels of the column factor.	$MS_{WG} = \frac{SS_{WG}}{df_{WG}} = \frac{\_\_\_\_\_\_}{\_\_\_\_} = \mathbf{48}$ $df_{WG} = df_{TOT} - df_{AG} = (N-1) - (k-1)$ = (_____ – 1) – (_____ – 1) = (_____ – _____) = **24** $MS_{Gender} = \frac{SS_{Gender}}{df_{Gender}} = \frac{\_\_\_\_\_\_}{\_\_\_\_} = \mathbf{1020.83}$ df_{Gender} = (r – 1) = (_____ – 1) = **1** $MS_{Dynamic} = \frac{SS_{Dynamic}}{df_{Dynamic}} = \frac{\_\_\_\_\_\_}{\_\_\_\_} = \mathbf{250}$ $df_{Dynamic}$ = (c – 1) = (_____ – 1) = **2** $MS_{Inter} = \frac{SS_{Inter}}{df_{Inter}} = \frac{\_\_\_\_\_\_}{\_\_\_\_} = \mathbf{626.98}$ $df_{Inter} = df_{Gender} * df_{Dynamic}$ = (_____ * _____) = **2** $F_{Gender\ dfGender,dfWG} = F_{Gender\_\_\_,\_\_\_} =$ $\frac{MS_{Gender}}{MS_{WG}} = \frac{\_\_\_\_\_\_}{\_\_\_\_}$ **= 21.267** $F_{Dynamic\ dfDynamic,dfWG} = F_{Dynamic\_\_\_,\_\_\_} =$ $\frac{MS_{Dynamic}}{MS_{WG}} = \frac{\_\_\_\_\_\_}{\_\_\_\_}$ **= 5.208** $F_{Inter\ dfInter,dfWG} = F_{Inter\_\_\_,\_\_\_} = \frac{MS_{Inter}}{MS_{WG}} = \frac{\_\_\_\_\_\_}{\_\_\_\_}$ **= 13.062**

Thought Questions

- **If H_O of gender is false (there *are* differences between genders), would you expect MS_{Gender} to be the same or larger than MS_{WG}? Explain how you know.**

STATISTICAL CONCLUSION	
Definitions	**Calculations**
The Gender Effect: This F-ratio tests the null hypothesis that there is no main effect of _______________.	$F_{Gender\_\_\_,\_\_\_}$ = **21.267** $_{.05}F_{crit\ \_\_\_,\_\_\_}$ = **4.26** Gender Main Effect Decision: **Reject** Fail to Reject
The Dynamic Effect: This F-ratio tests the null hypothesis that there is no main effect of _______________________.	$F_{Dynamic\_\_\_,\_\_\_}$ = **5.208** $_{.05}F_{crit\ \_\_\_,\_\_\_}$ = **3.40** Dynamic Main Effect Decision: **Reject** Fail to Reject
The Interaction Effect: This F-ratio tests the null hypothesis that there is no gender by group dynamic _______________ effect.	$F_{Inter\_\_\_,\_\_\_}$ = **13.062** $_{.05}F_{crit\ \_\_\_,\_\_\_}$ = **3.40** Interaction Effect Decision: **Reject** Fail to Reject

Graph of the means of each group:



Thought Question

- **What can you conclude about the task enjoyment of the people? Interpret your conclusion in terms of the main effects of gender, of dynamic, and of the interaction effect. The graph may help.**

- **Why could it be problematic to interpret the main effects of gender and dynamic (even if they were significant) when the interaction is significant?**

SUMMARY TABLE

Source of Variation	df	SS	MS	F	$_{.05}F_{crit}$	Decision
Within Groups	**24**					
Among Groups	**(5)**					
Gender					4.26	
Dynamic					3.40	
Gender x Dynamic					3.40	
TOTAL	**29**					

Stop Time: _______________

Name: ____

Section Time: **25 minutes for Parts Two & Three**

Part Three

Using purple pen, use Part 3 to do back and correct your work in Part 2:

3) Correct the blanks you filled in as needed, including the Summary Table

4) Edit or improve your answers to the thought questions using the answers that follow. Do not just copy the explanations over; it is important you modify or edit what you already have written. Be sure you include all the main ideas in the provided thought question answers.

Turn in Parts Two & Three to your TA to get the Quiz

Start Time: ____________

Two-Way ANOVA

In a study of group dynamics, 30 people were grouped into teams of 5, and were given directions for a group project that had them work cooperatively, neutrally, or competitively. There were 3 teams of women and 3 teams of men, each in one of the group dynamic conditions. This study has two factors: **Gender** (2 levels: women and men) and **Group Dynamic** (3 levels: cooperative, neutral, and competitive). They were then asked to rate their enjoyment of the project on a scale from 0-100, with 100 being highly enjoyable. Is there a difference in **task enjoyment** between **genders**? Between the different **dynamic** groups? Is there a **gender by dynamic interaction**?

THE DATASET

	Cooperative	Neutral	Competitive
Women	**95**	**75**	**35**
$\bar{X}_{Women}$ =	98	79	38
68.3	96	71	32
	84	66	26
	100	84	44
	97	75	35
Men	**40**	**50**	**80**
$\bar{X}_{Men}$ =	44	41	75
56.7	31	55	89
	36	50	80
	49	59	85
	40	45	71
$\bar{X}_{Total}$ = **62.5**	$\bar{X}_{Coop}$ = **67.5**	$\bar{X}_{Neut}$ = **62.5**	$\bar{X}_{Comp}$ = **57.5**

ΣX / T VALUES

	Cooperative	Neutral	Competitive	
Women	475	375	175	**1025**
Men	200	250	400 $_{Cells}$	**850** $_{Rows}$
	675	**625**	**575** $_{Columns}$	**1875** $_{Total}$
ΣX^2 = **133027**				

NULL HYPOTHESES

H_O Gender:	$\mu_{Women} = \mu_{Men}$
H_O Dynamic:	$\mu_{Coop} = \mu_{Neut} = \mu_{Comp}$
H_O Interaction:	*No Interaction*

SUMS OF SQUARES

Definitions	Shortcut Formula
$\mathbf{T_{group}} = \Sigma X_{group}$ **n** = # in Each Group **c** = # Columns **r** = # Rows The sum of squares **total** (**SS_{TOT}**) is a measure of the total amount of variability in the data. $SS_{TOT} = SS_{AC} + SS_{WG}$	$\mathbf{SS_{TOT}} = \Sigma X^2 - \frac{T^2_{Total}}{N_{Total}}$ $133027 - \frac{1875^2}{30}$ $= 133027 - 117187.5$ **= 15839.5**
The sum of squares **within groups** (**SS_{WG}**) is a measure of the amount of variability within each group due to noise. SS_{WG} = *noise*	$\mathbf{SS_{WG}} = \Sigma X^2 - \Sigma(\frac{T^2_{cell}}{n_{cell}})$ $133027 - \frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5}$ $= 133027 - 131875$ **= 1152**
The sum of squares **among** each of the 6 **groups** or cells (**SS_{AG}**) is a measure of the amount of variability due to the treatment effect and due to noise. *It is the same as* SS_{Cells}. SS_{AG} = ____________ *effect + noise* The SS_{AG} can be further partitioned into the SS of row, column, and interaction. $SS_{AG} = SS_{Gender} + SS_{Dynamic} + SS_{Interaction}$	$\mathbf{SS_{AG}} = \Sigma(\frac{T^2_{cell}}{n_{cell}}) - \frac{T^2_{Total}}{N_{Total}}$ $\frac{475^2}{5} + \frac{375^2}{5} + \frac{175^2}{5} + \frac{200^2}{5} + \frac{250^2}{5} + \frac{400^2}{5} - \frac{1875^2}{30}$ $= 131875 - 117187.5$ **= 14687.5**

The Sum of Squares of **Gender (SS_{Gender})** is the amount of variability across each of the 2 gender rows. *SS_{Gender} = effect of _______________ + noise* The Sum of Squares of Group **Dynamic ($SS_{Dynamic}$)** is the amount of variability across each of the 3 dynamic columns. *$SS_{Dynamic}$ = effect of _______________ + noise*	$\mathbf{SS_{Gender}} = \Sigma\left(\frac{T^2_{Gender}}{n_{Gender}}\right) - \frac{T^2_{Total}}{N_{Total}}$ $= \frac{1025^2}{15} + \frac{850^2}{15} - \frac{1875^2}{30}$ $= 118208.3 - 117187.5$ $= \mathbf{1020.83}$ $\mathbf{SS_{Dynamic}} = \Sigma\left(\frac{T^2_{Dynamic}}{n_{Dynamic}}\right) - \frac{T^2_{Total}}{N_{Total}} = \frac{675^2}{10} + \frac{625^2}{10} + \frac{575^2}{10} - \frac{1875^2}{30}$ $= 117687.5 - 117187.5$ $= \mathbf{500}$
The sum of squares of the **interaction ($SS_{Interaction}$)** is remaining variability among the groups that is not due to the effect of row and of column. Therefore, it must be due to the effect of the interaction between the row and column. *$\mathbf{SS_{Interaction}}$ = effect of _______________ + noise*	$\mathbf{SS_{Interaction}} = SS_{AG} - SS_{Gender} - SS_{Dynamic}$ $= 14687.5 - 1020.83 - 500$ $= \mathbf{13166.67}$

Thought Questions

- **The total variability is partitioned into each of these Sums of Squares (SSs). Describe how each SS listed above 'fits' together to form the SS_{TOTAL}.**

The SS total is made up of the SS within group and among group. The SS_{AG} is made up of the SS of the row, the column, and the interaction.

$$\mathbf{SS_{TOTAL} = SS_{WG} + SS_{AG}}$$

$$\mathbf{SS_{ROW} + SS_{COL} + SS_{INTERACT}}$$

MEAN SQUARES & F-RATIO

Definitions	Calculations
***k** = # Groups; **N** = Total # Datapoints* **c** = # Columns **r** = # Rows The Mean Square **(MS)** is the amount of variance within each partition. MS has the same formula as variance (S^2): $\mathbf{MS} = \frac{SS}{df} = \frac{SS}{n-1} = S^2$ It can be thought of as the *average* (or mean) variability per degree of freedom.	$\mathbf{MS_{WG}} = \frac{SS_{WG}}{df_{WG}} = \frac{1152}{24} = \mathbf{48}$ $\mathbf{df_{WG}} = df_{TOT} - df_{AG} = (N-1) - (k-1)$ $= (30-1) - (6-1) = (29-5) = \mathbf{24}$
	$\mathbf{MS_{Gender}} = \frac{SS_{Gender}}{df_{Gender}} = \frac{1020.83}{1} = \mathbf{1020.83}$ $\mathbf{df_{Gender}} = (r-1) = (2-1) = \mathbf{1}$
	$\mathbf{MS_{Dynamic}} = \frac{SS_{Dynamic}}{df_{Dynamic}} = \frac{500}{2} = \mathbf{250}$ $\mathbf{df_{Dynamic}} = (c-1) = (3-1) = \mathbf{2}$
	$\mathbf{MS_{Inter}} = \frac{SS_{Inter}}{df_{Inter}} = \frac{13166.67}{2} = \mathbf{626.98}$ $\mathbf{df_{Inter}} = df_{Gender} * df_{Dynamic} = (1 * 2) = \mathbf{2}$

The **F-ratio** puts the amount of variability among the rows, columns, or the interaction over the amount of variability within each group. **The Gender Effect:** Differential effects produced by different levels of the row factor. **The Dynamic Effect:** Differential effects produced by different levels of the column factor. **The Interaction Effect:** Differences that are produced by unique combinations of <u>gender</u> and <u>dynamic</u>. An interaction is present when the effects of the row factors change with different levels of the column factor.	$\mathbf{F_{Gender}}_{\ dfGender,dfWG} = F_{Gender\ 1,24}$ $= \frac{MS_{Gender}}{MS_{WG}} = \frac{1020.83}{48} = \mathbf{21.267}$ $\mathbf{F_{Dynamic}}_{\ dfDynamic,dfWG} = F_{Dynamic\ 2,24}$ $= \frac{MS_{Dynamic}}{MS_{WG}} = \frac{250}{48} = \mathbf{5.208}$ $\mathbf{F_{Inter}}_{\ dfInter,dfWG} = F_{Inter\ 2,24}$ $= \frac{MS_{Inter}}{MS_{WG}} = \frac{626.98}{48} = \mathbf{13.062}$

Thought Questions

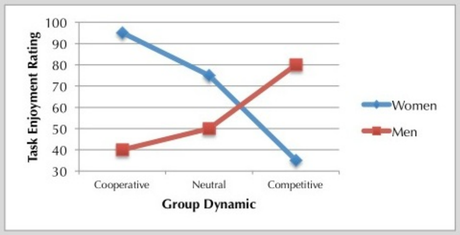
- **If H_O of gender is false (there *are* differences between genders), would you expect MS_{Gender} to be the same or larger than MS_{WG}? Explain how you know.**

If H_O about gender is false, MS_{Gender} should be larger than MS_{WG}. MS_{Gender} includes variability due to the effect of gender *and* noise. Since there is a gender effect, MS_{Gender} would include this *plus* the variability due to noise, whereas MS_{WG} only includes variability due to noise.

STATISTICAL CONCLUSION

Definitions	Calculations
The Gender Effect: This F-ratio tests the null hypothesis that there is no main effect of <u>gender</u>.	$\mathbf{F_{Gender\ 1,24}} = \mathbf{21.267}$ $\mathbf{_{.05}F_{crit\ 1,24}} = \mathbf{4.26}$ Gender Main Effect Decision: <u>**Reject**</u> Fail to Reject
The Dynamic Effect: This F-ratio tests the null hypothesis that there is no main effect of <u>dynamic</u>.	$\mathbf{F_{Dynamic\ 2,24}} = \mathbf{5.208}$ $\mathbf{_{.05}F_{crit\ 2,24}} = \mathbf{3.40}$ Dynamic Main Effect Decision: <u>**Reject**</u> Fail to Reject
The Interaction Effect: This F-ratio tests the null hypothesis that there is no gender by group dynamic <u>interaction</u> effect.	$\mathbf{F_{Inter\ 2,24}} = \mathbf{13.062}$ $\mathbf{_{.05}F_{crit\ 2,24}} = \mathbf{3.40}$ Interaction Effect Decision: <u>**Reject**</u> Fail to Reject

Graph of the means of each group:



Thought Question

- **What can you conclude about the task enjoyment of the people? Interpret your conclusion in terms of the main effects of gender, of dynamic, and of the interaction effect. The graph may help.**

There is an effect of **gender**, meaning there are differences in enjoyment between gender, regardless of dynamic. There is also an effect of **group dynamic**, meaning there are differences in enjoyment between dynamics, regardless of gender. There is also an **interaction** between gender and dynamic on enjoyment: the effect of dynamic depends on gender. Women preferred the cooperative dynamic most and disfavored the competitive; whereas the men preferred the competitive dynamic most and disfavored the cooperative.

- **Why could it be problematic to interpret the main effects of gender and dynamic (even if they were significant) when the interaction is significant?**

Men and women (regardless of dynamic) do differ in task enjoyment, and the three dynamic groups (regardless of gender) differ in task enjoyment as well. However, since there is an interaction, the effect is **not consistent**. This makes it difficult to say something conclusive about the individual main effects. For example, we cannot conclude that being cooperative was always preferred, because it was for women, but not for men.

SUMMARY TABLE

Source of Variation	df	SS	MS	F	$_{.05}F_{crit}$	Decision
Within Groups	**24**	1152	48			
Among Groups	**(5)**	(14687.5)				
Gender	1	1020.83	1020.83	21.267	$_{.05}F_{1,24} = 4.26$	Reject
Dynamic	2	500	250	5.208	$_{.05}F_{2,24} = 3.40$	Reject
Gender x Dynamic	2	13166.67	626.98	13.062	$_{.05}F_{2,24} = 3.40$	Reject
TOTAL	**29**	15839.5				

Stop Time: _______________

Name: ____

Quiz Time: **20 minutes**

Rating Items – 3 minutes

Quiz Items – 17 minutes or until end of class

Start Time: _______________

Section Time: **3 minutes**

Think about the calculations and thought questions you just completed. Please answer the following questions about your work with them in mind.
Please answer the following items on a scale from 1(low) to 5 (high), with 3 being the neutral point.

1)	In solving or studying these worksheets I invested _____ mental effort	Low	1	2	3	4	5	High
2)	How easy or difficult was it to complete these worksheets?	Easy	1	2	3	4	5	Difficult
3)	How mentally demanding were these worksheets?	Low	1	2	3	4	5	High
4)	How hurried or rushed was the pace of these worksheets?	Not Rushed	1	2	3	4	5	Very Rushed
5)	How successful were you in accomplishing what you were asked to do?	Not At All	1	2	3	4	5	Very
6)	How hard did you have to work to accomplish your level of performance?	Not At All	1	2	3	4	5	Very
7)	How insecure, discouraged, irritated, stressed, and annoyed were you?	Not At All	1	2	3	4	5	Very

For the following items, response to if you engaged in these activities either <u>in your head</u> or <u>on paper</u>, whether you were instructed to or not.

Use the following scale:

1- Rarely while working 2- Somewhat often while working 3- About half the time while working
4- Quite often while working 5- Almost always while working

		Rarely				Almost Always
14)	I thought deeply about what my calculated values represented and meant	1	2	3	4	5
16)	I copied the thought-questions exactly	1	2	3	4	5
17)	I paraphrased the thought-questions	1	2	3	4	5
19)	I explained in my own words what the thought-questions meant to me	1	2	3	4	5

20)	I developed my own answers to the thought-questions before reading the provided answers	1	2	3	4	5
22)	I connected the thought-questions to ideas I already knew	1	2	3	4	5
23)	I made hypotheses or predictions about the thought-questions	1	2	3	4	5
24)	I justified or provided reasons why concepts in the thought-questions occur	1	2	3	4	5

Stop Time: _______________

Start Time: _______________

Section Time: **17 minutes**

---

5) Which SS_{AG} is larger? SS_{AG} from Dataset A or Dataset B?

Dataset A:		**Dataset B:**	
H_O Row:	Reject	H_O Row:	Reject
H_O Column:	FTR	H_O Column:	Reject
H_O Interaction:	FTR	H_O Interaction:	FTR

Explain you reasoning:

---

2)

Table A	
Source	**SS**
Among Groups	
Age	572.93
Time	580.53
Age x Time	558.29
Within Groups	103.91
TOTAL	699.08

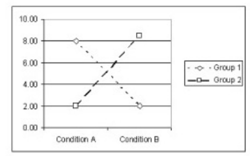
Table B	
Source	**SS**
Among Groups	
Age	590.85
Time	598.45
Age x Time	576.20
Within Groups	481.33
TOTAL	131.78

Table C	
Source	**SS**
Among Groups	
Age	5.63
Time	13.23
Age x Time	9.03
Within Groups	103.91
TOTAL	131.78

Which of the following tables has the correct calculations for the SSs? Circle One:

Table A Table B Table C

3) Interpret the following two-way ANOVA graphs by circling Yes or No:

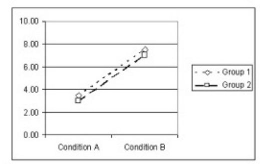


A)

Effect of Condition: Yes No

Effect of Group: Yes No

Interaction Effect: Yes No

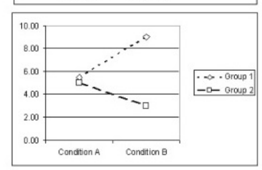


B)

Effect of Condition: Yes No

Effect of Group: Yes No

Interaction Effect: Yes No



C)

Effect of Condition: Yes No

Effect of Group: Yes No

Interaction Effect: Yes No

For Graph C, What can you conclude about the performance of the people? (Interpret your conclusion in terms of the main effects of condition, of group, and of the interaction effect.)

4) In a two-way ANOVA with 3 levels of Factor A and 4 levels of Factor B, with 7 subjects in each cell, what will be the value of the degrees of freedom within groups (df_{WG})?

a) 11
b) 72
c) 77
d) 84

What would be the degrees of freedom to use in looking up the ${}_{.05}F_{crit}$ for the interaction effect?

---

5) In a two-way ANOVA with 4 different groups, the smaller the values for the sample variances, the more likely it is that at least one of the F-ratios will be significant.

Var 1	Var 3
Var 2	Var 4

True   False

---

6) Complete the summary table and circle a decision for each:

Source of Variation	df	SS	MS	F	${}_{.05}F_{Crit}$	Decision
Among Groups						
Rows	1		10		4.17	
Columns		170		5.67	2.92	
Interaction			16.67		2.92	
Within Groups		320	10			
TOTAL		550				

Stop Time: ______________

Name: ____

1) In a study on city dwellers vs. country dwellers in the US and China, the city dweller mean was very different from the country dweller mean (regardless of country), and the US mean was about the same as the China mean (regardless of living environment).

What would you expect? Circle one for each:

$MS_{CityvsCountry}$ < > ≈ ? MS_{WG}

$MS_{USvsChina}$ < > ≈ ? MS_{WG}

$MS_{Interaction}$ < > ≈ ? MS_{WG}

2)

ΣX / T VALUES				
	Low	Med	High	
Young	40	60	80	180
Old	100	50	25 Cells	175 Rows
	140	110	105 Columns	355 Total
$\Sigma X^2 = 1500$				

There are 30 young and 30 old, with 10 in each cell.

Fill in the following formulas to show how you would determine them:

SS_{Age} : ** Do not have to calculate final value for this, just show how you would get there.*	**SS_{Age}** = $\qquad - \frac{T^2_{Total}}{N_{Total}}$
SS_{Level} :	**SS_{Level}** = $\qquad - \frac{T^2_{Total}}{N_{Total}}$

3) Complete the summary table and circle a decision for each.

There are 2 rows of condition and 4 columns of time, and 6 subjects in each cell, so 48 total. Start with the degrees of freedom:

Source of Variation	df	SS	MS	F	$_{.05}F_{Crit}$	Decision
Among Groups		280				
Condition					4.08	
Time			48		2.84	
Interaction		120			2.84	
Within Groups						
TOTAL		600				

Interpret your conclusion in terms of the main effects of condition, of time, and of the interaction effect.

Did you finish?
____ Yes ____ No
Would you have wanted more time?
____ Yes ____ No

APPENDIX B

TA Instructions

Time is TIGHT so please keep things moving briskly!

- Before class, write on board:
  - **Can get started now – No more than <u>3 mins</u> on Part 1**
  - **Write on name on each packet**
  - **Follow Directions Carefully**
  - **Legible handwriting**
  - **Write Start and Stop Times**
  - **Raise hand for next packet**

  - **Section Times:**
    - **Pretest: 3 minutes**
    - **Part 1 & 2: 25 minutes**
    - **Quiz: 20 minutes (till end of class)**
      - **Rank Items: 3 minutes**
      - **Quiz Items: 17 minutes**

- Put Clock up on computer in room for students to record their time:
  - http://www.online-stopwatch.com/large-digital-clock/
- Have Pretests already passed out on desks, or have students pick on up on their way in. As students come in, determine what color Part one they need to get. Do NOT allow students to look at Part one while they are working on the pretest.
- As people are getting settled in, they can be working. Since almost everyone will have do this last time, there's no need for big directions to the class. However, do announce:

<u>Announce to Class (once everyone is here):</u>

"Since there are only two sections before this exam, we want the Friday activities to really help you learn and understand the material. We are trying out some different versions of the worksheets to figure out what is most useful for you guys, but don't worry- they both have the exact same content.

These are going to be more challenging than what you're used to. Really try to do your best. Don't worry about grading, we know these are harder so we'll adjust our grading accordingly. The most important thing is to put in a good effort.

Time will be tight so it's really important to stay focused. There are a bunch of different sections, and we have time limits on each section. This is just so we're sure to get through all the material, and to be sure to give you enough time for the quiz.

There will be a pretest, two parts and then the quiz. I've noted the amount of time you have for each part on the board. If you finish early, you can go ahead to the next section after checking with me. Keep your eye on the time and try to pace yourself so you get through everything, but if you can't get to everything, that's totally fine. I'd rather have you do half of it well than all of it rushed.

- In front of your is the pretst and Part 1. Work through the pretest when I say 'start'.
- When you are done, raise your hand. I'll collect the pretest, and then you can get started on Part 1
- When you are done with Part 1, raise your hand and I will give you Part 2
- When you are done with Part 2, hand in Parts 1 and 2 to me, and I will give you the quiz.

The first part of the quiz has some questions that ask about how much effort you put in on the worksheet. That part is not for a grade, so please be honest. It's important for us to know exactly how you were thinking about the material (even if it's not very much!) so we can figure out how to make the worksheets more useful for you.

The quiz items might be challenging. Just do your best.

<u>A couple things to remember:</u>

- Write your name on the front page of every packet
- It is super important you read the directions very carefully, and follow them as closely as you can
- On the computer screen is a clock. On the packets, there will be places to jot down your Start and Stop times. Be sure you do this.
- This is supposed to be a little challenging, don't worry about grading. The most important thing is to put in a good effort.
- There will be 3 parts and then the quiz. I've noted the amount of time you have for each part on the board. When you're done, raise your hand and I'll give you the next part.
- If you finish early, you can go ahead to the next section after checking with me. Keep your eye on the time and try to pace yourself so you get through everything, but if you can't get to everything, that's totally fine.

Alright, go ahead and get started!!

<u>Notes for you:</u>

Don't give much help. If a student has a question, just tell them to re-read the problem carefully, and do their best.

If Low on Time: Can have student skip the "On a scale from 1-5" page (3 mins)

APPENDIX C

Item Screening Values

Assessment Item	Flag - ged	Item-Total Correlation	N	Mean	Std Dev	Infit	Outfit	Item Difficulty	Item Difficulty SE	Scale Alpha	Alpha If Deleted
Week One											
Pretest											
W1_P1_1_Trans_Valid_C		0.42	74	0.51	0.50	0.87	0.7	-1.06	0.31	0.624	0.551
W1_P1_1_Trans_HQ_C		0.46	74	0.16	0.37	0.78	0.45	1.78	0.41	0.624	0.546
W1_P1_2_Trans_VALID_P	*	0.32	74	0.81	0.39	1.28	6.57	-4.43	0.62	0.624	0.592
W1_P1_2_Trans_HQ_P		0.40	74	0.41	0.49	0.92	0.76	-0.3	0.31	0.624	0.564
W1_P1_3_Trans_Valid_C		0.27	74	0.30	0.46	1.18	1.01	0.49	0.33	0.624	0.615
W1_P1_3_Trans_HQ_C	*	0.31	74	0.07	0.25	1.09	0.68	3.52	0.65	0.624	0.603
Posttest											
W1_P4_Quiz_1_HO_Score_C		0.24	74	0.73	0.45	1.51	1.69	-1.35	0.33	0.858	0.860
W1_P4_Quiz_1_MSAG_Score_P		0.40	74	0.86	0.34	1.09	1.19	-2.66	0.42	0.858	0.854
W1_P4_Quiz_1_Why_Valid_C		0.45	74	0.62	0.49	1.13	1.03	-0.58	0.3	0.858	0.852
W1_P4_Quiz_1_Why_HQ_C		0.33	74	0.27	0.45	0.99	1.08	1.58	0.3	0.858	0.856
W1_P4_Quiz_2_Size_Score_C		0.47	74	0.85	0.36	0.92	1.15	-2.49	0.4	0.858	0.851
W1_P4_Quiz_2_Why_Valid_C		0.58	74	0.85	0.36	0.76	0.62	-2.49	0.4	0.858	0.848
W1_P4_Quiz_2_Why_HQ_C		0.32	74	0.24	0.43	1.05	0.9	1.76	0.31	0.858	0.857
W1_P4_Quiz_3_Decision_Score_P		0.51	74	0.72	0.45	1.06	0.78	-1.25	0.32	0.858	0.849
W1_P4_Quiz_3_dfAG_Score_P		0.48	74	0.81	0.39	1.06	0.74	-2.06	0.36	0.858	0.851
W1_P4_Quiz_3_dfWG_Score_P		0.42	74	0.77	0.42	1.14	1.3	-1.69	0.34	0.858	0.853
W1_P4_Quiz_3_F_Score_P		0.71	74	0.58	0.50	0.66	0.55	-0.32	0.29	0.858	0.840
W1_P4_Quiz_3_MSWG_Score_P		0.74	74	0.62	0.49	0.61	0.5	-0.58	0.3	0.858	0.839
W1_P4_Quiz_3_SSAG_Score_P		0.62	74	0.58	0.50	0.82	0.72	-0.32	0.29	0.858	0.844
W1_P4_Quiz_3_SSTot_Score_P		0.64	74	0.50	0.50	0.73	0.66	0.18	0.29	0.858	0.843
W1_P4_Quiz_4_Score_C		0.44	74	0.88	0.33	0.97	1.07	-2.84	0.44	0.858	0.853
W1_P4_Quiz_5_Valid_C		0.38	74	0.74	0.44	1.22	1.38	-1.46	0.33	0.858	0.854
W1_P4_Quiz_5_HQ_C		0.51	74	0.32	0.47	0.79	0.6	1.23	0.29	0.858	0.849
W1_P4_Quiz_6_F_Score_C		0.40	74	0.55	0.50	1.2	1.15	-0.15	0.29	0.858	0.854
W1_P4_Quiz_6_MSAG_Score_C		0.15	74	0.11	0.31	1.19	1.06	2.93	0.4	0.858	0.860
W1_P4_Quiz_6_SSAG_Score_C		0.15	74	0.12	0.33	1.23	1.1	2.78	0.38	0.858	0.861
W1_P4_Quiz_7_Valid_C	*D	0.10	74	0.03	0.16	1.04	0.7	4.52	0.73	0.858	0.860
W1_P4_Quiz_7_HQ_C	*D	0.06	74	0.01	0.12	1.03	0.86	5.25	1.01	0.858	0.860
Retention											
W1_P5_1_InDe_Score_C		0.59	74	0.65	0.48	1.12	0.96	-2.43	0.35	0.935	0.932
W1_P5_1_InDeWhy_Valid_C		0.64	74	0.46	0.50	0.97	0.8	-0.93	0.32	0.935	0.932
W1_P5_1_InDeWhy_HQ_C		0.52	74	0.34	0.48	1.2	1.28	-0.02	0.32	0.935	0.933
W1_P5_2_Trans_Valid_C	*	0.42	74	0.55	0.50	1.5	2.01	-1.64	0.32	0.935	0.935
W1_P5_2_Trans_HQ_C		0.31	74	0.35	0.48	1.67	3.79	-0.13	0.32	0.935	0.936
W1_P5_3_dfAG_Score_P		0.67	74	0.61	0.49	0.93	0.73	-2.08	0.34	0.935	0.931
W1_P5_3_dfWG_Score_P		0.70	74	0.58	0.50	0.84	0.58	-1.85	0.33	0.935	0.931
W1_P5_3_HO_Score_P		0.61	74	0.64	0.48	1.1	0.87	-2.31	0.35	0.935	0.932
W1_P5_3_MSAG_Score_P		0.77	74	0.50	0.50	0.65	0.48	-1.23	0.32	0.935	0.930
W1_P5_3_SSTot_Score_P		0.79	74	0.36	0.48	0.56	0.41	-0.23	0.32	0.935	0.930
W1_P5_3_SSWG_Score_P		0.83	74	0.42	0.50	0.49	0.37	-0.63	0.32	0.935	0.929
W1_P5_3_Why_Valid_C		0.71	74	0.61	0.49	0.78	0.51	-2.08	0.34	0.935	0.931
W1_P5_3_Why_HQ_C		0.64	74	0.27	0.45	0.86	0.66	0.52	0.34	0.935	0.932
W1_P5_4_Decision_Score_P		0.42	74	0.16	0.37	1.06	1.4	1.55	0.39	0.935	0.934
W1_P5_4_F_Valid_P		0.56	74	0.41	0.49	1.15	1.05	-0.53	0.32	0.935	0.933
W1_P5_4_F_HQ_P	*D	0.43	74	0.05	0.23	0.61	0.16	3.29	0.6	0.935	0.934
W1_P5_4_SSAG_Valid_P		0.45	74	0.19	0.39	1.13	1.14	1.26	0.37	0.935	0.934
W1_P5_4_SSAG_HQ_P	*D	0.50	74	0.08	0.27	0.65	0.21	2.69	0.5	0.935	0.934
W1_P5_4_Fcritdf_Valid_P		0.59	74	0.36	0.48	1.04	1.17	-0.23	0.32	0.935	0.932
W1_P5_4_Fcritdf_HQ_P	*D	0.37	74	0.11	0.31	1.22	0.78	2.24	0.45	0.935	0.935
W1_P5_4_MSAG_Valid_P		0.68	74	0.41	0.49	0.86	0.68	-0.53	0.32	0.935	0.931
W1_P5_4_MSAG_HQ_P	*D	0.51	74	0.08	0.27	0.61	0.2	2.69	0.5	0.935	0.934
W1_P5_4_MSWG_Valid_P		0.71	74	0.38	0.49	0.78	0.59	-0.33	0.32	0.935	0.931
W1_P5_4_MSWG_HQ_P	*D	0.43	74	0.05	0.23	0.61	0.16	3.29	0.6	0.935	0.934
W1_P5_4_SSTot_Valid_P	*	0.49	74	0.66	0.48	1.18	2.85	-2.56	0.36	0.935	0.934
W1_P5_4_SSTot_HQ_P	*D	0.36	74	0.35	0.48	1.58	2.07	-0.13	0.32	0.935	0.936
W1_P5_4_SSWG_Valid_P		0.53	74	0.30	0.46	1.05	1.43	0.3	0.33	0.935	0.933
W1_P5_4_SSWG_HQ_P	*D	0.55	74	0.12	0.33	0.69	0.33	2.05	0.43	0.935	0.933

Week Two											
Pretest											
W2_P1_1_Valid_C		0.60	74	0.70	0.46	0.98	0.45	-3.03	0.48	0.767	0.709
W2_P1_1_Quality_C		0.45	74	0.20	0.40	1.03	0.67	2.45	0.41	0.767	0.749
W2_P1_2_Valid_P		0.60	74	0.76	0.43	0.76	0.18	-4.14	0.59	0.767	0.710
W2_P1_2_Quality_P		0.46	74	0.36	0.48	1.17	1.25	0.76	0.36	0.767	0.747
W2_P1_3_Valid_C		0.51	74	0.39	0.49	1.05	1.82	0.51	0.36	0.767	0.734
W2_P1_3_Quality_C		0.47	74	0.14	0.34	0.75	0.3	3.45	0.49	0.767	0.746
Posttest											
W2_P4_Quiz_1_Score_C		0.62	74	0.78	0.41	0.97	0.93	-2.59	0.44	0.927	0.924
W2_P4_Quiz_1_Why_Valid_C		0.55	74	0.68	0.47	1.08	1.21	-1.47	0.33	0.927	0.925
W2_P4_Quiz_1_Why_Quality_C		0.47	74	0.38	0.49	1.05	1.55	0.5	0.29	0.927	0.926
W2_P4_Quiz_2_Score_C		0.54	74	0.68	0.47	1.11	1.41	-1.47	0.33	0.927	0.925
W2_P4_Quiz_3_ACond_Score_C	*D	-0.13	74	0.11	0.31	1.54	5.03	2.64	0.41	0.927	0.931
W2_P4_Quiz_3_AGroup_Score_C	*D	-0.01	74	0.04	0.20	1.07	5.67	3.83	0.61	0.927	0.929
W2_P4_Quiz_3_AInter_Score_C	*D	0.40	74	0.73	0.45	1.27	3.49	-1.95	0.37	0.927	0.927
W2_P4_Quiz_3_BCond_Score_C		0.39	74	0.53	0.50	1.27	1.43	-0.43	0.29	0.927	0.927
W2_P4_Quiz_3_BGroup_Score_C	*	0.37	74	0.59	0.49	1.34	2.03	-0.87	0.3	0.927	0.927
W2_P4_Quiz_3_BInter_Score_C		0.62	74	0.76	0.43	0.91	0.91	-2.24	0.4	0.927	0.924
W2_P4_Quiz_3_C_Interpret_Valid_C		0.64	74	0.73	0.45	0.84	0.83	-1.95	0.37	0.927	0.924
W2_P4_Quiz_3_C_Interpret_Quality_C		0.37	74	0.24	0.43	1.1	1.09	1.4	0.32	0.927	0.927
W2_P4_Quiz_3_CCond_Score_C	*	0.08	74	0.12	0.33	1.36	2.06	2.48	0.39	0.927	0.929
W2_P4_Quiz_3_CGroup_Score_C		0.61	74	0.74	0.44	0.96	0.83	-2.09	0.38	0.927	0.924
W2_P4_Quiz_3_CInter_Score_C		0.41	74	0.55	0.50	1.21	1.36	-0.6	0.3	0.927	0.927
W2_P4_Quiz_4_Score_P		0.35	74	0.53	0.50	1.4	1.5	-0.43	0.29	0.927	0.927
W2_P4_Quiz_4_df_Score_P		0.29	74	0.14	0.34	1.04	1.26	2.34	0.38	0.927	0.927
W2_P4_Quiz_5_Score_C		0.45	74	0.57	0.50	1.24	1.29	-0.69	0.3	0.927	0.926
W2_P4_Quiz_6_DesCol_Score_P		0.50	74	0.58	0.50	1.09	1.24	-0.78	0.3	0.927	0.925
W2_P4_Quiz_6_DesInter_Score_P		0.74	74	0.39	0.49	0.59	0.52	0.41	0.29	0.927	0.922
W2_P4_Quiz_6_DesRow_Score_P		0.69	74	0.39	0.49	0.69	0.6	0.41	0.29	0.927	0.923
W2_P4_Quiz_6_dfCol_Score_P		0.76	74	0.34	0.48	0.51	0.42	0.75	0.3	0.927	0.922
W2_P4_Quiz_6_dfInter_Score_P		0.67	74	0.42	0.50	0.79	0.7	0.24	0.29	0.927	0.923
W2_P4_Quiz_6_dfTot_Score_P		0.69	74	0.28	0.45	0.6	0.47	1.12	0.31	0.927	0.923
W2_P4_Quiz_6_dfWG_Score_P		0.65	74	0.45	0.50	0.82	0.73	0.08	0.29	0.927	0.923
W2_P4_Quiz_6_FInter_Score_P		0.76	74	0.36	0.48	0.53	0.45	0.58	0.29	0.927	0.922
W2_P4_Quiz_6_FRow_Score_P		0.76	74	0.39	0.49	0.56	0.48	0.41	0.29	0.927	0.922
W2_P4_Quiz_6_MSCol_Score_P		0.75	74	0.34	0.48	0.53	0.44	0.75	0.3	0.927	0.922
W2_P4_Quiz_6_SSInter_Score_P		0.65	74	0.43	0.50	0.83	0.74	0.16	0.29	0.927	0.923
W2_P4_Quiz_6_SSRows_Scores_P		0.68	74	0.54	0.50	0.83	0.69	-0.52	0.29	0.927	0.923
Retention											
W2_P5_1_CC_Score_C	*D	0.34	74	0.53	0.50	1.53	3.16	-2	0.32	0.942	0.944
W2_P5_1_Inter_Score_C	*D	-0.09	74	0.12	0.33	2.97	9.9	2.1	0.48	0.942	0.947
W2_P5_1_USC_Score_C	*D	0.03	74	0.03	0.16	1.41	5.03	4.42	0.78	0.942	0.944
W2_P5_2_SSAge_Valid_P		0.52	74	0.58	0.50	1.09	1.09	-2.41	0.33	0.942	0.941
W2_P5_2_SSAge_Quality_P		0.46	74	0.24	0.43	1.74	1.49	0.42	0.39	0.942	0.942
W2_P5_2_SSLevel_Valid_P		0.55	74	0.49	0.50	1.19	1.03	-1.69	0.32	0.942	0.941
W2_P5_2_SSLevel_Quality_P		0.48	74	0.22	0.41	1.65	1.54	0.74	0.41	0.942	0.941
W2_P5_3_DesCond_Score_P		0.81	74	0.18	0.38	0.42	0.2	1.27	0.44	0.942	0.938
W2_P5_3_DesInter_Score_P		0.69	74	0.16	0.37	0.78	0.52	1.47	0.45	0.942	0.939
W2_P5_3_DesTime_Score_P		0.65	74	0.15	0.36	0.83	0.59	1.67	0.46	0.942	0.939
W2_P5_3_dfCond_Score_P		0.55	74	0.68	0.47	0.92	0.65	-3.27	0.38	0.942	0.941
W2_P5_3_dfInter_Score_P		0.62	74	0.47	0.50	0.9	0.94	-1.59	0.32	0.942	0.940
W2_P5_3_dfTime_Score_P		0.57	74	0.65	0.48	0.87	0.6	-3	0.36	0.942	0.940
W2_P5_3_dfTot_Score_P		0.57	74	0.45	0.50	1.14	1.17	-1.39	0.32	0.942	0.940
W2_P5_3_dfWG_Score_P		0.71	74	0.38	0.49	0.8	0.68	-0.86	0.33	0.942	0.938
W2_P5_3_FCond_Score_P		0.84	74	0.18	0.38	0.33	0.16	1.27	0.44	0.942	0.937
W2_P5_3_FInter_Score_P		0.79	74	0.16	0.37	0.47	0.2	1.47	0.45	0.942	0.938
W2_P5_3_FTime_Score_P		0.85	74	0.22	0.41	0.41	0.23	0.74	0.41	0.942	0.937
W2_P5_3_Inter_Valid_C	*D	0.69	74	0.22	0.41	0.91	0.88	0.74	0.41	0.942	0.939
W2_P5_3_Inter_Quality_C	*D	0.00	74	0.00	0.00	1	1	6.49	1.84	0.942	0.944
W2_P5_3_MSCond_Score_P		0.83	74	0.23	0.42	0.49	0.28	0.58	0.4	0.942	0.937
W2_P5_3_MSInter_Score_P		0.72	74	0.31	0.47	0.86	0.7	-0.27	0.36	0.942	0.938
W2_P5_3_MSWG_Score_P		0.84	74	0.24	0.43	0.5	0.3	0.42	0.39	0.942	0.937
W2_P5_3_SSCond_Score_P		0.83	74	0.24	0.43	0.54	0.32	0.42	0.39	0.942	0.937
W2_P5_3_SSTime_Score_P		0.75	74	0.35	0.48	0.74	0.57	-0.63	0.34	0.942	0.938
W2_P5_3_SSWG_Score_P		0.69	74	0.35	0.48	0.91	1.11	-0.63	0.34	0.942	0.939