AN EXPERIMENTAL STUDY COMPARING TWO DIFFERENT TRAINING STRATEGIES ON HOW TO USE STATISTICAL PACKAGES IN AN INTRODUCTORY STATISTICS COURSE

James Baglin, Cliff Da Costa, Matthew Ovens, Vicky Bablas

Presenting Author: James Baglin (james.baglin@rmit.edu.au)
School of Mathematics and Geospatial Sciences, RMIT University, Melbourne VIC 3001, Australia

KEYWORDS: statistics education, statistical packages, error management training, educational experiment

ABSTRACT
Statistics education reform advocates the use of statistical software packages, such as *SPSS*, *Minitab*, and *R*, in introductory statistics courses. However, little research has focused on finding effective methods of training students how to use these packages. This experimental study evaluated the effectiveness of two different training strategies. A sample of 100 first year psychology students were randomly allocated to either a *Guided* training (GT) strategy or an *error management training* (EMT) strategy. GT consisted of step-by-step instructions that aimed to explicitly direct participants through training to use the statistical package *SPSS*. EMT avoided step-by-step instructions, instead focusing on minimal instruction which aimed to engage participants in actively exploring the statistical package. In EMT, errors were viewed as beneficial to training as they help develop a deeper understanding of *SPSS*. Participants completed five statistical package training sessions across the semester. After four of these training sessions, participants completed self-assessment tasks that measured training transfer performance. After controlling for statistical knowledge and self-assessment compliance, the results of the study indicated no statistically significant differences between the two training strategies. However, this conclusion must be interpreted with caution due to a number of study imitations which are discussed in this article.

Proceedings of the Australian Conference on Science and Mathematics Education, University of Melbourne, Sept 28th to Sept 30th, 2011, pages 162-168, ISBN Number 978-0-9871834-0-8.

INTRODUCTION
Technology has had a great impact on educational practices in statistics education over the last couple of decades. In 2005, the American Statistics Association stated in their Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report (American Statistical Association, 2005) that 'Almost any course in statistics can be improved by more emphasis on data and concepts, at the expense of less theory and fewer recipes. To the maximum extent feasible, calculations and graphics should be automated' (p. 4). This recommendation set the stage for the widespread deployment of technology, especially statistical packages, into statistics courses (Garfield, Hogg, Schau, & Whittinghill, 2002). Fortunately there is a plethora of statistical packages available to statistics instructors with common examples which include *SPSS*, *Minitab*, and *R*. The main advantages offered to statistics instructors by statistical packages involve automating complex statistical formulae, developing tools for demonstrating statistical concepts and giving students valuable experience with practical quantitative research software.

Despite the widespread adoption of statistical packages in statistics courses, very little is known about how students learn to use these packages. Anecdotally, most students demonstrate poor recall of operational knowledge between week-to-week computer tutorials. The poor recall is further compounded between semesters. This suggests that current strategies of training students to use statistical packages may be ineffective.

Guided methods are probably the most common training strategies employed in statistics courses (Mills, 2003). *Guided* training (GT) is based on the programmed learning method developed by Skinner (1968). According to GT the learner is a passive participant during training (Keith, Richter, & Naumann, 2010). Step-by-step, comprehensive and explicit instructions aim to guide the learner through the features and procedures of the statistical package. The GT method views errors made during training as a waste of time which need to be avoided. Mastery of the package comes through repeated practice where operational errors are avoided. GT in statistics courses can be delivered in a number of ways including the use of commercially published books (e.g. Allen & Bennett, 2008; Francis, 2007), course print-outs, or labs developed for online learning systems. Online learning systems have the advantage of embedding instructions, tasks, questions, and feedback into the one system. The major disadvantage of GT courseware is that it can be very time consuming to develop and maintain, especially with the frequent updating of statistical package software. Research also suggests that there are other equally or even more effective training strategies available (Keith & Frese, 2008).

A recent meta-analysis of 24 studies found that GT for computer software (e.g. word processors, email and spreadsheets) was significantly less effective compared to an alternate form of training known as *error management training* (EMT, Keith & Frese, 2008). EMT is a type of *active/exploratory* learning (Bell & Kozlowski, 2008) which uses minimal information to engage trainees in active exploration of a software package (Frese, Brodbeck, Heinbikel, Mooser, Schleiffenbaum, & Thiemann, 1991). Thus, students are assumed to be active participants in the training process (Bell & Kozlowski, 2008). Minimal instruction encourages errors which EMT frames in a positive way. Errors are argued to be beneficial to training as they promote exploration, help develop the know-how to avoid errors and the know-how to overcome errors once they have been committed (Frese, et al., 1991). Overall, errors help develop a much deeper understanding of a software package that might otherwise be missed in GT. EMT frames errors in a positive light by presenting heuristics to students during training such as “Errors are a natural part of learning. They point out what you can still learn!" (Dormann & Frese, 1994, p. 368). This positive attitude towards errors is further reinforced by instructors during training. The idea underlying the positive error framing is to promote exploration and also help students regulate their negative emotions towards making errors during training.

Research suggests that EMT is superior to GT for training transfer (Keith & Frese, 2008). Training transfer can be defined as knowledge and skills gained during training which transfer to other tasks and jobs outside of training (Hesketh, 1997). The literature differentiates between two major types of training transfer. *Analogical transfer* relates to tasks that are similar to those covered during training (Keith & Frese, 2008; Keith, et al., 2010), whilst on the other hand, *adaptive transfer* tasks consists of tasks that are structurally distinct from training tasks and require the trainee to adapt their knowledge gained from training in novel ways (Ivancic & Hesketh, 1996; Keith, et al., 2010). Adaptive transfer is considered to be a more desirable outcome to training due to the relative shortness of training. It is important that statistical package training gives a student the foundational knowledge and skills to build upon and use in later courses.

The meta-analysis study by Keith and Frese (2008) found that EMT was moderately more effective when compared with GT for analogical transfer, but was substantially more effective for adaptive transfer. As Keith et al. (2010) explains, active/exploratory type training, such as EMT, works by developing a learner’s metacognitive skills which involve planning, monitoring and evaluating task performance (Brown, Bransford, Ferrara, & Campione, 1983). In contrast, students taught with GT are not given the opportunity to develop these self-regulatory skills. GT is setup to avoid the need for metacognitive skills above and beyond what is needed to follow step-by-step instructions (Ivancic & Hesketh, 1996). When students are placed in a novel situation outside of training, students in the EMT training are able to apply the metacognitive skills they developed and practiced in training, whereas students from GT have difficulty because they have not developed these metacognitive skills needed to generalise from the training environment (Keith, et al., 2010). Whilst there is substantial support for the effectiveness of EMT for general software training, very little research has focused specifically on statistical package training.

In one experiment by Dormann and Frese (1994), EMT was compared with GT for training psychology students to use the statistical package *SPSS*. Dormann and Frese found that the EMT group performed significantly better than the GT group on both moderate and difficult transfer tasks. Despite these positive findings in favour of EMT, there were a number of limitations to the study. The training outcome was measured immediately after the training session. Week-to-week or long-term training transfer retention was not measured. Also, Dormann and Frese’s study was a well-controlled experiment done outside of a regular statistics course. No study has assessed the ecological validity of EMT in a real introductory statistics course filled with all the regular challenges of low student motivation, training non-compliance, statistical anxiety, training apathy and student stress.

Therefore, the aim of this study was to investigate the ecological effectiveness of EMT for learning to use a statistical package in an introductory statistics course. It was hypothesised that EMT would be comparable to GT for analogical transfer tasks, but that EMT would be superior to GT for adaptive transfer tasks.

METHOD

PARTICIPANTS

Participants consisted of 1st year psychology students enrolled in an introductory statistics course which ran concurrently across two campuses. Students were randomly assigned to odd and even week computer labs as part of a regular course requirement. Of the 151 students enrolled, 117 consented to participate in the experiment. Three of these consenting students were not randomly allocated but instead placed automatically into unfilled labs due to space limitations. There were 14 consenting participants who did not finish training. The following table displays the characteristics of the sample across the EMT and GT strategies.

Table 1: Sample Characteristics of Groups

		Strategy		
		GT	EMT	Total
Strategy	N (%)	44 (44%)	56 (56.0%)	100
Campus A	N (%)	10 (33.3%)	20 (66.7%)	30
Campus B	N (%)	34 (48.6%)	36 (51.4%)	70
Female	N (%)	31 (44.3%)	39 (55.7%)	70
Male	N (%)	13 (43.3%)	17 (56.7%)	30
Age	M ± SD	19.84 ± 5.02	19.41 ± 5.07	19.60 ± 5.02

MEASURES

Training outcome measures consisted of two self-assessment tasks that were completed in the final weeks of training between Lab session 4 and 5. The first self-assessment tasks consisted of 8 exercises that measured a student's analogical transfer. These exercises were similar to tasks that students had completed during training. The second self-assessment task originally consisted of 8 exercises that measured adaptive transfer, but 4 of these exercises were eliminated due to an incorrect data file being supplied which resulted in the EMT groups not being able to attempt the last four exercises. Adaptive transfer tasks were structurally distinct from training and required students to complete tasks and analyses in *SPSS* that were not strictly covered during training. A total transfer score was also computed by summing together analogical and adaptive transfer scores.

Students were given 25 minutes to complete each self-assessment task. However, as students were able to complete labs outside of allocated lab times, many students went over this time limit. Students were instructed that to get the grade for the self-assessment, they would need to get 4/8 on the first assessment and 2/4 on self-assessment 2. Questions were randomised from pools of similar questions. Participants were allowed to attempt each self-assessment up to 5 times as the labs and self-assessment were graded on completion (formative assessment). Only a participant's first attempt score was recorded for analysis.

When designing the self-assessment tasks, it was important that each task measured a student's ability to successfully operate the statistical package and not be confounded by the student's knowledge of statistics. For example, completing an exercise task that gets a student to find the median IQ of the sample may be confounded by the student's knowledge of the median. While it was virtually impossible to eliminate this statistical knowledge dependency, each exercise task was designed to minimise its effect. For example, exercise questions which were used to score someone on their ability to operate *SPSS* asked questions relating to the acquired output from *SPSS* that proved they had completed the analysis correctly. The questions avoided interpretation of statistics or graphs which would be dependent on student's statistical knowledge. Regardless, statistical knowledge, which was defined as the proportion of marks obtained on the end of year multiple-choice exam, was included as a covariate in the statistical analysis of the results. It was assumed that the statistics exam score was an indicator of both the academic ability and statistical knowledge of the student.

Compliance on the self-assessment was also taken into consideration. Compliance was defined as whether the student completed both self-assessment tasks in the allocated self-assessment lab. If the students completed any of the self-assessment tasks outside of the allocated self-assessment lab, they were classified as non-compliant.

PROCEDURE

The method and the results reported in this paper form the first semester, primary outcomes of an ongoing project looking at how students learn to use statistical packages. The entire ongoing project covers two semesters. Following university ethics approval and random allocation to odd and even week computer labs, students were approached before their lecture to participate in the study. Non-consenting students were still required to complete training, but their data was not processed for the study. The allocation to different labs was due to limitations with size and availability of large computer labs. This odd and even week group allocation allowed for the manipulation of training strategies. The ordering of EMT and GT to odd and even weeks was counterbalanced between the Campuses. Campus A had GT on odd weeks and EMT training on even weeks. On campus B the order was reversed. Counterbalancing the order controlled for possible time effects introduced by using odd and even weeks.

Training consisted of five labs covering the following: An introduction to *SPSS*, The Basics of *SPSS*, Frequencies in *SPSS*, Cross Tabulation in *SPSS* and Correlation and Regression in *SPSS.* Self-assessment was completed towards the end of the semester between Lab 4 and 5. Labs were scheduled for one hour per week, however students were permitted to stay longer to finish or catch-up. Students who missed their designated labs needed to ask permission to attend a non-designated lab. This was done so as to not disadvantage students and was a condition for ethics approval. This meant that some students were mixing strategies. This issue and the fact that students talk to each other, made blinding participants to the strategies impossible. However, the exact nature of the strategies was never explained to participants.

Training was delivered using a streamlined, proprietary, online assessment system called WebLearn. WebLearn is similar to a streamlined version of Blackboard's© quiz, test and assignment features. Each lab consisted of objectives, instructions and exercises embedded with the strategies' instructions. The labs coincided with content delivered in lectures. The GT group received step-by-step comprehensive instructions and screen shots summarising each exercise in *SPSS*. The students were instructed to follow these steps and answer questions that confirmed they had completed the exercise correctly. GT was designed to minimise errors during training. The EMT strategy was given the exact same exercises but with modified instructions and no screen shots. The EMT training used minimal instruction to get the participant actively exploring *SPSS*. Instructions were designed to point the participants in the right direction, but students were left to work out the specifics. Sometimes for difficult analyses, hints were given to help students get back on track if they got stuck. Students were also presented with error management heuristics listed at the top of each exercise. Examples of these heuristics included "If you have a problem, regard it as a learning opportunity" (Wood, Kakebeeke, Debowski, & Frese, 2000) and "Errors are a natural part of learning. They point out what you can still learn". These heuristics aimed to help participants frame errors in a positive way so that they could learn from them.

A lab supervisor was also present at each scheduled lab. In the GT strategy, the supervisor was instructed to help the participants as much as they needed in line with the theory of GT. In the EMT strategy, the supervisor was advised to encourage the students to find the solution themselves. If the participant was struggling after multiple attempts, the supervisor was allowed to give them a hint to get them back on track. The supervisor was also trained to reinforce the positive error framing by encouraging students to learn from their mistakes.

RESULTS

Descriptive statistics for training transfer outcomes and covariates are contained in Table 2. The EMT strategy was more compliant with self-assessment, but had lower statistically knowledge when compared to the GT strategy. In terms of training transfer, the EMT strategy outscored the GT strategy on analogical transfer, but not for adaptive transfer.

One-way analysis of covariance (ANCOVA) was performed to assess for significant differences between the GT and EMT groups on measures of adaptive transfer, analogical transfer and total transfer (see Table 3). The ANCOVA used self-assessment compliance and statistical knowledge as covariates. Table 3 contains the ANCOVA model parameters and covariate adjusted means for all three training transfer outcomes. The ANCOVA for analogical transfer and adaptive transfer were not statistically significant with $F(1,96) = 2.34$, $p = 0.129$ and $F(1,95) = 0.122$, $p = 0.73$ respectively. Total training transfer was also not statistically significant $F(1,96) = 2.17$, $p = 0.144$.

Table 2: Descriptive Statistics for Training Transfer Measures and Covariates

		Strategy	
		GT	EMT
Self-assessment Compliance	Yes N (%)	21 (47.7%)	40 (71.4%)
	No N (%)	23 (52.3%)	16 (28.6%)
Statistical Knowledge	M ± SD	0.74 ± 0.15	0.69 ± 0.15
Analogical Transfer Score	M ± SD	5.23 ± 1.92	5.41 ± 1.66
Adaptive Transfer Score	M ± SD	1.72 ± 0.88	1.66 ± 0.94
Total Transfer Score	M ± SD	6.91 ± 2.31	7.07 ± 2.16

In all models, statistical knowledge was a statistically significant covariate which indicated that there was a positive relationship between training transfer and statistical knowledge (Table 3). Compliance was also a statistically significant covariate for analogical transfer and total transfer, but not for adaptive transfer. According to the ANCOVA models in Table 3, non-compliance was associated with higher analogical and total transfer scores.

Table 3: ANCOVA Models Predicting Training Transfer Measures

Parameters	Analogical		Adaptive		Total Transfer	
	B	95% CI	B	95% CI	B	95% CI
Self-assessment Compliance[a]	-0.92	(-1.62, -0.22)	-0.38	(-0.77, 0.001)	-1.25	(-2.12, -0.39)
Statistical Knowledge	4.17	(1.90, 6.44)	1.55	(0.31, 2.79)	5.64	(2.84, 8.44)
Strategy[b]	-0.53	(-1.22, 0.16)	-0.07	(-0.45, 0.31)	-0.63	(-1.48, 0.22)
GT Adjusted Mean (95% CI)[c]	5.05	(4.54, 5.57)	1.66	(1.38, 1.94)	6.68	(6.04, 7.31)
EMT Adjusted Mean (95% CI)[c]	5.59	(5.15, 6.02)	1.73	(1.49,1.96)	7.31	(6.77, 7.85)

[a] Compliant students = 1 [b] Reference category was EMT [c] Means after adjusting for the covariates of compliance and statistical knowledge

DISCUSSION

The results of this study found no statistically significant differences in training transfer between training strategies after controlling for statistical knowledge and self-assessment compliance. These results fail to support the research hypotheses and conflict with the previous findings of Dormann and Frese (1994) which found that EMT was significantly better than GT. However, there are a number of limitations in this study that must be considered before drawing conclusions.

In terms of the study design, this experiment was un-blinded. While students were never explicitly made aware of the nature of this study, it is highly probable that students became aware of the difference between conditions as the semester progressed. The supervisor was also un-blinded to the nature of the conditions. While it is difficult to speculate the exact influence this might have had on results, future research needs to blind participants and supervisors to control for potential bias. This will be a difficult an experiment of this kind.

The major strength of this study, ecological validity, was also its greatest limitation. Due to many issues with access to large computer labs, training was scheduled on a fortnightly basis for each group. This meant that students had only a minimum estimated training time of 4 hours with SPSS before taking the self-assessment tasks. Given the large time intervals between training and the relative shortness of training, it is possible that the effects of training were interrupted and poorly consolidated.

The training labs were compulsory, but a large number of students missed labs on a regular basis. Due to ethical reasons, these students were permitted to attend labs of the opposite condition or complete the labs in their own time. However, these students still received their respective condition's instructions as the labs were delivered through an online learning system which based lab availability on their allocated condition. As attendance was recorded at all labs, it is possible for future analyses of results to include adherence to allocated lab conditions as a covariate in the modelling of training transfer performance. This would not only control for poor training adherence, but also the potential bias of un-blinded participants.

The labs were scheduled for only one hour. While the training was designed to fit within this time period, many students reported feeling time pressure which resulted in them rushing through labs and using guesswork to get the lab done in the designated time. It is possible that time constraints negatively impacted the EMT and violated the error framing instructions. Under time constraints, it would be very difficult for a student to view errors as anything else but a waste of time. While the availability of computer labs was outside the control of the researchers, a possible solution to this problem would be to get students working outside of class in their own time. That way at least they would not be under the time pressure and more inclined to frame errors in a positive sense.

An error with self-assessment two, adaptive transfer tasks resulted in 4 out of 8 questions being removed from the analysis. This error only affected the EMT group. It is difficult to determine what would have happened if the issue did not occur, but it would be safe to assume that the inclusion of four more adaptive transfer tasks would have introduced more variability in adaptive transfer scores and made it easier to detect differences between strategies if those differences existed.

All training was graded in terms of satisfactory completion and students were allowed multiple attempts at the training labs and self-assessment tasks. This feature of training may have resulted in unmotivated students not expending their greatest effort on self-assessment tasks. Instead, they may have done just enough to attain a level of satisfactory completion. The issues of low incentive may have masked a participant's true ability on the self-assessment tasks.

In conclusion, the results of this study found no statistically significant difference between GT and EMT for training to use a statistical software package. However, further research is needed to overcome the limitations of this study and shed more light on strategies for learning to use statistical packages. It remains to be seen whether "less is more" when it comes to training students how to use statistical packages in introductory statistics courses.

ACKNOWLEDGEMENTS

This study was approved by the RMIT College Human Ethics Advisory Network on the 6th December 2010 (Application No: BSEHAPP 48 – 10).

REFERENCES

Allen, P., & Bennett, K. (2008). *SPSS for the health and behavioural sciences*. South Melbourne, VIC: Thomson.

American Statistical Association (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) project*: Author.

Bell, B. S., & Kozlowski, S. W. J. (2008). Active learning: Effects of core training design elements on self-regulatory processes, learning, and adaptability. *Journal of Applied Psychology, 93*, 296-316.

Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Handbook of child psychology. In P. Mussen, J. H. Flavell & E. M. Markman (Eds.), (Vol. 3, pp. 77-166): Harvard University Press.

Dormann, T., & Frese, M. (1994). Error training: Replication and the function of exploratory behavior. *International Journal of Human-Computer Interaction, 6*, 365-372.

Francis, G. (2007). *Introduction to SPSS for Windows: v15.0 and 14.0 with notes for Studentware*. Frenchs Forest, NSW: Pearson Education Australia.

Frese, M., Brodbeck, F., Heinbokel, T., Mooser, C., Schleiffenbaum, E., & Thiemann, P. (1991). Errors in training computer skills: On the positive function of errors. *Human-Computer Interaction, 6*, 77--93.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education, 10*(2), 456--467.

Hesketh, B. (1997). Dilemmas in training for transfer and retention. *Applied Psychology, 46*, 317-339.

Ivancic, K., & Hesketh, B. (1996). Making the best of errors during training. *Training Research Journal, 1*, 103-125.

Keith, N., & Frese, M. (2008). Effectiveness of error management training: A meta-analysis. *Journal of Applied Psychology, 93*, 59-69.

Keith, N., Richter, T., & Naumann, J. (2010). Active/exploratory training promotes transfer even in learners with low motivation and cognitive ability. *Applied Psychology: An International Review, 59*, 97-123.

Mills, J. D. (2003). SPSS textbooks: A review for teachers. *Statistics Education Research Journal, 2*, 59-70.

Skinner, B. F. (1968). *The technology of teaching*. New York: Appleton-Century-Crofts.

Wood, R., Kakebeeke, B., Debowski, S., & Frese, M. (2000). The impact of enactive exploration on intrinsic motivation, strategy, and performance in electronic search. *Applied Psychology, 49*, 263-283.