Batanero, C. (Ed.), Training Researchers in the Use of Statistics, 231-244.

Printed in Granada, Spain.

CAROL JOYCE BLUMBERG

TRAINING REGULAR EDUCATION AND SPECIAL EDUCATION TEACHERS IN THE USE OF RESEARCH METHODOLOGY AND STATISTICS

The purpose of this paper is to discuss the needs of primary and secondary level teachers seeking Master's degrees (second degrees) in terms of knowledge and use of research methodology and statistics, with special consideration of the needs of Special Education teachers. In particular, the following topics will be discussed: goals and organisation; descriptive statistics; inferential statistics; specific hypothesis testing procedures; experimental and quasi-experimental designs; survey research and sampling techniques; qualitative data collection techniques; and reliability and validity. The paper ends with a discussion of what special education teachers need to know concerning meta-analysis and single-subject designs and with some miscellaneous comments.

1. INTRODUCTION

The purpose of this paper is to discuss the needs of primary and secondary level teachers seeking Master's degrees (that is, second degrees) in terms of knowledge and use of research methodology and statistics, with special consideration of the needs of special education teachers. In the USA, the term special education teacher refers to any teacher who has taken extra training (usually as part of a first degree) to be able to teach students with disabilities, such as mental retardation (the exact term used actually varies from state to state in the USA), learning disabilities, emotional problems, and physical disabilities. The term research methodology and statistics is used here to mean the traditional statistical methods, as well the related areas of statistical thinking, logic of hypothesis testing, experimental and quasi-experimental design, survey research design, sampling, qualitative data collection techniques, reliability and validity.

To the best of the author's knowledge, almost all Master's degree programs in Education or Special Education in the USA and England include such courses. In this paper the discussion will be in terms of all teachers who are pursuing second degrees in Education or Special Education, except where otherwise noted. However, since there is a special emphasis in this paper on Special Education teachers, many of the references will be in terms of special education even though other, more general, references were sometimes available.

2. TEACHERS AS RESEARCHERS

While it may seem strange at first to label a classroom teacher also as a researcher, it is not. All teachers are continually collecting and analysing data on students' academic

performance. Special Education teachers, besides collecting and analysing data on academic performance, also must perform well-designed experiments in their classrooms with individual students or small groups of students (most often using single-subject design principles) to assess what "treatments" will be most effective for educating each student (Tawney & Gast, 1990). This is especially important in countries where, by law, each student with a disability must have an individual education plan (often abbreviated, IEP in the United States).

3. GOALS AND ORGANIZATION OF RESEARCH METHODOLOGY AND STATISTICS COURSES

The two major goals of most research methodology and/or statistics courses for teachers completing a Master's degree are to: i) help them critically read, correctly interpret, and decide the validity of conclusions in the published and unpublished literature, and ii) give them the tools necessary to complete research-based projects and/or theses required for their Master's degrees (Cooke, Test, Heward, Spooner, & Courson, 1993; Gall, Borg, & Gall, 1996). For Special Education teachers, there is a third major goal; they should learn how to better collect and analyse data that will be used to help determine each student's IEP. These three goals are exactly the same as those stated by Burrows and Baillie (1997) for the research methodology component of a Diploma in Nursing course in the UK with the phrase "help determine each student's IEP" above replaced by "enhance the quality of care delivered to the client" (Burrows & Baillie, 1997, p. 35).

Only one study of the content of the research methodology and/or statistics courses for Master's degree students in education could be found in the literature. Todd and Reece (1990) conducted a Delphi study on graduate introductory research courses. At the first stage, "...21 experts nominated by... [their peers] generated a list of 114 skills and knowledge areas they considered worthy objectives in an introductory course (p. 2-3). At the second stage, these 21 experts simplified the list to 59 skills and knowledge areas that are "'Essential', 'Essential-Important', or 'Important.'"(p. 3), where the authors had operational definitions for each of these terms.

This list of 59 skills and knowledge areas was made up of phrases, most of which were vaguely stated. For example, two of the areas were "Demonstrates an understanding of research methodology" and "Understands the concept of internal validity" (p. 6). Only a few were specific such as "Can formulate a testable hypothesis" (p. 6) and "Has an understanding of the standard deviation" (p. 7). This present paper, on the other hand, will provide more detailed guidelines as to how a faculty member might approach the content of research methodology and statistics courses for teachers. Also, reference will not be made in the rest of this paper to the number of experts who stated that certain skills and knowledge areas should be included, since doing so would be misleading, because so few of the specific (that is, not vaguely stated) skill and knowledge areas discussed here were included in Todd and Reece's (1990) list.

One major question that the faculty members in each Master's Degree program need to address in the beginning when designing the research methodology and statistics components of their program is whether or not research methodology and statistics should be combined in one course or split into two courses. Each has its advantages and disadvantages. When doing the literature search for this paper, only one article could be found that gave statistics on the number of courses required by various Master's Degree

programs. That article (Calder, Justen, &Waldrop, 1986) was a 1985 survey of 60 Master's Degree programs in Special Education. They found that 6 (10%) of these programs required no research or statistics courses, 29 (48.3%) required one course, 19 (31.7%) required two courses and 6 (10%) required three courses. For those programs that required only one course, they did not specify whether this was a research methodology course, statistics course or some combination of both. They did say that for the programs requiring two courses that most programs had one course each in statistics and in research methodology. No discussion could be found in the literature of the advantages and disadvantages of the various course combinations. So, the advantages and disadvantages listed here are the opinions of the author of this paper based on 20 years of experience in teaching research methodology and statistics.

The advantage of having research methodology and statistics combined into one course is that it is more efficient in terms of delivery of information. That is, unnecessary overlap between courses is eliminated. A second advantage of having one course is that the students are less intimidated before the course by the thought of taking a statistics course. The disadvantage of having the two combined is that is extremely hard to find faculty to teach such a course who are experts in both the research methodology used in education and in statistics. The advantage of separating research methodology and statistics into two courses is that it is easier to find faculty capable of teaching the courses. It is also true that that by separating the course, there will be overlap in the material covered. Although others may look at this as a disadvantage, the author of this paper feels that it is an advantage as long as the faculty members teaching the two courses communicate well with each other to minimise needless overlap.

4. DESCRIPTIVE STATISTICS

Before discussing the specifics of what teachers pursuing Master's degrees need to know about descriptive statistics and other topics, it should be pointed out that in the USA few first-degree (Bachelor's) programs require students preparing to be teachers to take a course in statistics. Hence, for this paper the assumption is made that the students enrolled in a Master's degree course in research methodology and statistics have no knowledge of statistics and probability other than how to compute the mean, median, and mode.

It is extremely important that several hours are spent on the proper use and interpretation of descriptive statistics since these techniques are so widely used in the literature, as well as in the technical portions of standardised test manuals. Elmore and Woehlke (1998) have pointed out, that out of the 1906 articles examined that were published in the three journals of the American Educational Research Association (AERA) from 1978 to 1997, graphical methods were used in 58 (3.0%) and descriptive statistics were used in 246 (12.9%) of the articles. In fact, descriptive statistics were the second most common technique used, with only the use of ANOVA, ANCOVA, and MANOVA combined being more common. It should be noted here that these percentages are artificially low since Elmore and Woehlke used all articles published as their population, whether or not the article used any methods of data collection or data reporting. Baumberger and Bangert (1996) found the same order, but because of the way they reported their results no exact percentages can be given here. Heckenlaible-Gotto and Choi (1993) found that 34% of the 94 articles they examined from the 1990 to 1993 issues of *Topics in Early Childhood Special Education* used descriptive

statistics. Those interested in the percentages of use of statistical techniques in the biological and health sciences are encouraged to read the paper by Harraway, Manly, Sutherland, and McRae (2001).

Graphics: Some discussion needs to be done of how to present readable (in terms of axes and legends) histograms or bar charts of both frequencies and relative frequencies. Also, since boxplots and stem-and-leaf displays are increasing every year in their occurrence in journal articles and in textbooks for students of ages 10 and up, it is important that these teachers know how to interpret both boxplots and stem-and-leaf displays. They should also learn how to make a stem-and-leaf displays by hand. It was two Special Education teachers who mentioned to the author that they had been asked for help with how to construct a stem-and-leaf display by Special Education students who were mainstreamed for mathematics. Neither was able to help the students and both urged the author to include the construction of stem-and-leaf displays in her course. The term mainstreaming refers to when a special education student is taught mathematics (or some other subject matter area) in a regular classroom with either only a regular teacher present or with both a regular teacher and Special Education teacher present.

Measures of central tendency: Some discussion of when each of the measures of central tendency (that is, the mean, trimmed mean, median, and mode) should be used needs to take place. Most teachers can compute a mean, median, and mode. But, few have any knowledge as to when each should and should not be used.

Measures of variation & sampling distributions: Although most teachers regularly see standard deviations and standard errors, they do not know how to compute or interpret them. They are especially confused by the term standard error, even though it occurs in almost all standardised test manuals. So, the idea of a sampling distribution needs to be introduced to help them understand how a standard deviation and standard error are different. They also need to understand the difference between sampling error and other types of errors in data collection.

Hand/computer calculations: As many students in many different courses have pointed out to the author over the years, sometimes you learn some statistics concepts better by doing them step-by-step by hand/calculator first and then learning how to do them on a computer or using the special keys on calculators. Hence, it seems important that teachers be taught to do stem-and-leaf displays, histograms, means, medians, modes, variances, and standard deviation step-by-step by hand and/or by calculator. The inclusion of means here may seem strange, but there have been students in the author's Master's level research methodology class over the years that did not know how to compute the mean both by hand and by a calculator.

Use of computers: Since the computer packages available to teachers at their places of employment or at home change quickly, not much time should be spent training them in the use of a specific computer package. On the other hand, it is important that they have some exposure doing the descriptive statistics discussed in this section of this paper on a computer using a statistics package, so that they can begin to understand the logic of statistics packages. Further, the examination of output from at least two different statistics packages is important so that teachers realise how different the output may be from one package to another.

5. INFERENTIAL STATISTICS

The logic of hypothesis testing: In order to be able to critically read and correctly

interpret research-based articles, teachers must understand the logic of hypothesis testing, including the ideas of null hypothesis versus alternative hypothesis, of population parameter versus sample statistic, and of critical region including the correct use of the terminology of "do not reject the null" and "reject the null". Extensive discussion of the difference between statistical significance and practical (also called, meaningful or educational) significance also needs to take place.

Writing hypotheses: It is extremely important that teachers learn how to write both null and alternative hypotheses that are well defined (that is, operationally defined). By learning how to write hypotheses where all terms are operationally defined, they can better figure out what data needs to be collected and can get more useful help from a statistician before they collect data.

Probability values: There must be a careful discussion of probability values in these courses, since almost all published articles that use hypothesis tests in education present p-values. It is not necessary to have the teachers learn how to calculate probability values. They must, however, be able to correctly interpret probability values and be able to read them off of computer outputs.

Type I error, Type II error, and power: The distinction between Type I and Type II errors needs to be discussed. The concept of power should be introduced informally. They must understand conceptually how sample size, differences between means, and variability affect power and how power and Type II errors are related.

Confidence intervals: Teachers, like other disciplines at the Master's degree level, need to be able to correctly read and interpret confidence intervals. They must understand the correct way to write out and verbally state the interpretation of a confidence interval in language that parents can understand. They should also realise, as with t-tests, there are many different types of confidence intervals and one must be very careful as to what parameter or difference in parameters in being estimated by a confidence interval. In addition, because teachers deal with students as individuals, they must understand informally the difference between confidence intervals on means versus confidence (or prediction) intervals for an individual student's test scores or other measures of interest. The related topic of standard error of measurement also needs to be discussed.

Generalisation: In applied research there are two types of generalisations. The first is from the sample to the population that it represents. This is what inferential statistics does. The second type is the logical extension of statistically significant results from the population that the sample represents to other populations. This second type of generalisation is actually the one done more often in education. It is important that teachers learn to appreciate that both types of generalisation are acceptable as long as the people doing the generalisation make it clear which type they are doing.

Consulting with a statistician: It should be made clear to these teachers that the best way for them to complete their theses or other research papers required for their degree is to consult with a statistician before they do any inferential statistics. In fact, they should be strongly urged to consult with a statistician when they design their studies, before they begin data collection, as changes are being made during data collection, and before beginning their data analysis.

6. SPECIFIC HYPOTHESIS TESTING PROCEDURES

This is a big question in the author's mind as to how much these teachers need to

know about specific hypothesis testing procedures. For all of these procedures they should not have to compute anything by hand/calculator or by computer, but rather be able to interpret results.

Analysis of variance/analysis of covariance: Some discussion of how to read the results from ANOVAs, MANOVAs and ANCOVAs as reported in journal articles is absolutely necessary since these are the techniques used the most often according to Elmore and Woehlke (1998) and Baumberger and Bangert (1996). In fact, 309 (16.2%) articles out of 1906 articles appearing between 1978 and 1997 in the three AERA journals (Elmore & Woehlke, 1998) used these methods. ANOVA was also the most often used primary statistical technique in the special education literature in the period of 1984-85, being used in 27% of the random sample of 104 articles that Swanson and Alford (1987) examined. Teachers taking a research methodology and/or statistics course should have some conceptual understanding of the ideas of Sum of Squares, Mean Squares, and F-ratios. They should also be shown how to get the Sum of Squares, Mean Squares, degrees of freedom, F-ratios, and p-values from the output generated by one or two computer packages. The author's preference is for two computer packages here that present the results of an ANOVA differently.

It is also very important to make sure that teachers understand the ideas of main effects versus interactions and the ideas of covariates. They should understand the importance of correctly interpreting statistically significant interactions since interactions are extremely common in educational settings. For example, Torgesen and Dice (1980) found that in 62% of studies where the student (or subject) by experimental condition interaction was examined, this interaction was found to be statistically significant. Some discussion of how to interpret statistically significant main effects when there are also statistically significant interactions present should also take place, although probably only through the use of graphical displays of the appropriate means for the main effects and the interactions.

Correlation, regression, and contingency table analysis: Correlations and multiple regression are a very popular statistics in education, ranking as the fifth and third most used techniques, respectively, in articles published from 1978 to 1997 in the three journals of the AERA (Elmore & Woehlke, 1998). Contingency table analysis (that is, chi-square tests) was the primary statistical technique used in 7.7% of the articles examined by Swanson and Alford (1987). Correlation is also extensively used when studying reliability and validity of standardised tests. Hence, students must be able to interpret a correlation in terms of direction (positive versus negative) as well as in terms of practical significance and statistical significance. Often, R^2 is reported in the education literature. Hence, some discussion of how to interpret R^2 needs to be done as well as a discussion of the relationship between correlation and R^2. Time must also be spent helping these teachers understand how to interpret the results of regression and of contingency table analyses as reported in journal articles. They should also understand some of the major advantages and shortcomings of correlation, regression, and contingency table analysis. As Estepa and Sanchez Cobo (2001) have pointed out, the interpretation and correct use of measures of association is a difficult one for students taking a first course in statistics.

T-tests: Some discussion must take place about the various types of t-tests so that these teachers know to read and interpret the results carefully for t-tests appearing in research articles. The first several times the author taught a research methodology course to teachers she avoided introducing formulas. However, one time several members of the class asked to see the formula for a t-test. The author then showed them

the formula for an independent samples pooled t-test. It was amazing how much they learned from seeing the formula for the test statistic. Hence, some discussion of the formulas for a one-sample t-test and an independent samples t-test should be part of the course. Some explanation should also be included about how the various pieces of each of the formulas relate to the power of the test.

7. EXPERIMENTAL AND QUASI-EXPERIMENTAL DESIGN

Most of the projects and theses that Master's degree students complete have fairly simple designs. Hence time should not be spent teaching them a catalogue of complicated designs. They should also be exposed to the more common true experimental designs used in educational research: the Pretest-Posttest Control Group, the Posttest-Only Control Group, and the Solomon Four-Group (Gall, Borg, & Gall, 1996).

Teachers also need to be aware of the importance and usefulness of multi-factor designs. They also need to know the difference between random sampling and random assignment and why both are important. Further, time should be spent explaining why both random sampling and random assignment are hard to achieve in educational settings. This is especially true in special education because of legal and ethical requirements. For example, the USA (Public Law (P.L.) 94-142--The Education for All Handicapped Children Act), Israel (Michael, 1989), and Italy (Italian Law 517, 8/04/1977 and Italian Law 104, 2/05/1992) have laws that require special education students to be educated in the least restrictive environment possible and be placed in special classrooms only when their needs cannot be met in regular classrooms. The references for the Italian laws are from Balboni and Pedrabissi (2000). Hence, most well-designed research studies in special education settings (and even many in regular educational settings) are quasi-experimental. For example, Swanson and Alford's (1987) examination of 179 research articles in special education for the period of 1984-85 found that 0% were true experimental, 68% were quasi-experimental, and 32% were other types (for example, survey research or case studies). Hence, extensive discussion needs to take place on how to best design research studies using quasi-experimental methods that minimise bias as much as possible.

The proper analysis of quasi-experimental designs is still a matter of much debate among statisticians and can be very tricky (Cook & Campbell, 1979; Elashoff, 1969; Maxwell & Delaney, 1999). So, it is especially important that teachers doing research using quasi-experimental designs be urged to consult a statistician before they collect their data, if they need to make any changes while collecting their data, and before they begin their data analysis.

In education, the terms correlational studies and causal-comparative studies are often used. The term correlational study does not usually mean a study that is analysed using correlation. It is used to designate a study where one can only argue a relationship between two or more variables, but cannot argue that changes in one variable caused changes in another variable. Causal-comparative studies, on the other hand, are those where one can argue that changes in one variable caused changes in another variable. The distinction between these two terms, and why this distinction is important, needs to be made clear to these teachers.

8. SURVEY RESEARCH AND SAMPLING TECHNIQUES

Because of their perceived simplicity, surveys are a very popular instrument used in Master's degree research papers and theses in education as well as in schools and classrooms. For example, of the 19 theses and professional papers done by Master's degree students at Winona State University to which the author had access (see Appendix for a list), surveys were a major method of data collection in ten of them. Swanson and Alford (1987) found that survey research or other similar descriptive techniques were the basis for 18% of the articles they examined. Bruininks, Wolman, and Thurlow (1990) have pointed out that almost all follow-up studies of special education service programs include survey research. A follow-up study is one in which those who have graduated from and/or dropped-out of a special education program are contacted some number of years later and asked a series of questions. The answers to these questions are often the core of reports to local school districts, state agencies, and federal agencies.

Also, since almost all of these teachers will design a survey at sometime in their professional life it is important that substantial time be spent on how to write surveys that are clear and concise and collect useful information. Extensive discussion needs to take place of the various formats for items on surveys such as rating scales, semantic differentials, multiple choice, and open-ended items. Some discussion of the number of alternatives presented on a rating scale needs to take place. For example, should it be a 3-, 4-, 5-, 6-, or 7-point scale and should the respondents be allowed to choose neutral as a response? The issue of why it is essential to collect accurate background demographic information also needs to be discussed.

The related issue of confidentiality needs to be discussed, although classroom teachers are much more aware of the issues of confidentiality than any other researchers the author has worked with over the last 25 years. Further, they need to be made aware of the different ways of collecting data via mail, telephone, in-person, and the computer (both via email and the Internet). Some discussion of the merits and shortcomings of each method in a few different situations also needs to take place.

In addition these teachers need to be taught how to correctly use a random number table either in a book or on a computer. They also need to be shown good examples of the different sampling methods of simple random sampling, systematic random sampling, stratification, and clustering. Some discussion of how to make good strata and clusters also needs to be done. They should also see an example where a combination of sampling techniques was employed.

9. QUALITATIVE DATA COLLECTION TECHNIQUES

Time must be spent showing teachers a variety of qualitative data collection techniques. In fact, qualitative techniques were the fourth most popular techniques used in articles published in AERA journals between 1978 and 1997 (Elmore & Woehlke, 1998). Another reason these techniques should be discussed is that it is important for teachers to think more broadly when designing their research studies.

It is especially important that observational techniques be discussed since observational studies are extremely common in education and in special education, in particular (Greenwood, Peterson, & Sideridis, 1994-95). In fact, the use of observational techniques predates the use of quantitative techniques in special education (Hulek,

1983). Some of the issues that need to be discussed regarding observational techniques are minimal-bias data collection, the variety of recording and data-coding procedures (such as frequency-count recording and interval recording) available and the difficulty of properly analysing observational data.

Techniques such as case studies, cross-case analysis, ethnography, evaluation, and interviews also need to be discussed since, along with observational studies, they are the most popular qualitative techniques used in special education research (Crowley, 1994-95). Some general discussion of data coding and analysis under these techniques needs to occur. It needs to be emphasised that when data is collected using observational or other qualitative techniques, the issue of data analysis is a tough one. Further, it should be explained that often a variety of qualitative and quantitative techniques need to be used together to properly collect data (McWilliam, 1991) that will answer the research questions and/or hypotheses of interest.

10. RELIABILITY AND VALIDITY

Depending on the requirements of the individual State, Province, or Country, teachers may have had from very little exposure to extensive exposure to the topics of the reliability and validity of tests and other data collection instruments when pursuing their first degrees and initial teacher certification. Fairly typically in the States of Minnesota and Wisconsin (where the teachers the author has as students come from), there is little to no training in reliability and validity. It is important for these teachers to learn about reliability and validity for two reasons. First, they will be able to better understand the technical portions of the standardised test manuals. Second, when designing their own research projects they will be able to think better about these two topics. It is amazing how often teachers can make their studies much more useful by thinking about the reliability and validity consequences of the instruments they plan to use before collecting their data.

In terms of content here, they need to see at a minimum the ideas of internal consistency reliability (Cronbach's coefficient α), parallel forms reliability, reliability over time (often called, test-retest reliability), inter-rater reliability, content validity, construct validity, and criterion-related (both concurrent and predictive) validity. It also needs to be emphasised that these terms refer to specific measurements, while the terms internal validity and external validity refer to research studies as a whole.

11.ADDITIONAL TOPICS THAT NEED TO BE COVERED IN COURSES FOR SPECIAL EDUCATION TEACHERS

Meta-analysis: Meta-analysis has become a very popular technique in Special Education journals because of the small sample sizes that are inherent in research studies in special education due to the rarity of certain disabilities, such as autism or the combination of severe hearing and visual impairments or for other reasons. It should be remembered that meta-analysis was popularised first by educational statisticians (Glass, McGaw & Smith, 1981; Hedges & Olkin, 1985) and then spread to other areas of inquiry. As early as 1984 and 1985-86, the *Journal of Special Education* (1984; 1985-86) devoted two issues to meta-analysis and the more general issue of research synthesis. Stanovich and Stanovich (1997) have pointed out meta-analysis is also very

important in the area of Special Education because special education teachers have more faith in a meta-analysis of studies than in individual studies. Hence, Special Education teachers must be able to understand the logic of meta-analysis, have some knowledge of its advantages and disadvantages, and be able to interpret effect sizes in the simplest cases.

Single-subject design: In the USA there is a wide variety in the backgrounds of Special Education teachers in terms of knowledge of single-subject designs from having had an entire course in single-subject design prior to being certified as a Special Education teacher to having never even heard of the term, with most teachers having heard of single-subject design and having been taught how to use a few simple designs in their classrooms (Cooke, Test, Heward, Spooner, & Courson, 1993). In this course, a wide variety of single-subject designs need to be introduced along with how to implement and analyse them properly (Tawney & Gast, 1990). At a minimum, ABA (where A is the first treatment (usually the control treatment) and B is the second treatment), ABAB, other alternating treatment designs, and multiple baseline designs need to be introduced since these are the most popular single-subject designs used in educational research (Gall, Borg, & Gall, 1996). The teachers also need to understand how these designs can be used in rigorously done research as well as in the classroom.

Some discussion of how many observations need to be taken at each phase and how to determine the time intervals for collecting observations should be included. Further, it should be emphasised that the term single-subject design is a bit misleading. Some of the most useful single-subject studies are those where the same single-subject design is implemented on several students and the results discussed both individually and as a group (Tawney & Gast, 1990). Finally, some discussion of the limitations of single-subject design needs to take place including the limitations of the usual method of simply using a visual analysis. Since there has been much discussion over the last 30 years, at least, of how to statistically analyse single-subject designs (see, for example, Kratochwill & Levin, 1992), the topic of the statistical analysis of single-subject designs should probably not be discussed.

12. MISCELLANEOUS COMMENTS

Statistical thinking: The author does not see this as a separate topic in a research methodology or statistics course for teachers. Rather, it underlies the teaching of all of the material. Students often comment that after they have taken a Master's level research methodology or statistics course (the author has taught a Master's level statistics course for nursing many times), they think differently. When they explain what they mean by this it sounds like what is now called by the term statistical thinking.

Bias, external validity, and internal validity: These ideas should be discussed early and often throughout the course. As was pointed out several times in this paper, rarely is true experimental research possible in education and special education. Hence, the research which teachers read and do will contain various types of bias and internal validity problems. These biases and internal validity problems will limit the external validity (that is, the ability to generalise) of these studies. On the other hand, some generalisation can often be done and these teachers need to realise that just because a study contains bias or internal validity problems it is not useless. When reading studies they have to learn how to recognise bias problems and to decide whether or not the biases are too severe for the study to be of use to them. When doing their own studies,

they should try to eliminate as many bias problems as possible. These teachers must also be taught to be honest when reporting their results and to make clear how their sampling was done and admit to any other bias problems of which they are aware.

Replication: Some discussion needs to take place of the good points of doing replication studies. In fact, the only bad point of doing a replication study in the past was that many journals and university promotion and tenure committees in education would not accept them as being worthwhile. The situation is changing, however, as Gersten, Lloyd and Baker with others (1998) have pointed out. The good points of doing replication studies at the Master's degree level is that the teacher has a model to follow closely and hence the design of the study is made easier. Second, by having Master's degree students and others do replication studies, the quality of meta-analyses and other research syntheses will be improved because more studies will be available that can be compared and contrasted (Stanovich & Stanovich, 1997).

Other statistics topics: There are other statistics topics that have not been discussed in this paper because they occur less often in studies in education and special education. For the topics of factor analysis and structural equation modelling, the author has examples from journal articles that she distributes to her class and gives them a quick (10 minute) explanation of how to read the tables and diagrams for each technique. When the students in the class encounter other techniques in articles that they are reviewing as part of the course requirements, the author uses that article (or another one that is clearer if she has one) to explain a very little bit about the technique. Some of the statistical techniques that are used occasionally in special education are, in alphabetical order, ARIMA modelling and other time-series techniques, Bayesian analyses, cluster analysis, discriminant analysis, generalizability theory, log-linear models, multidimensional scaling, non-parametric and distribution-free techniques other than chi-square contingency table analysis (with, as mentioned in Section 6, chi-square contingency table analysis being a popular technique), small space analysis and survival analysis.

Training of first-degree (Bachelor's) students

As mentioned earlier, in the USA, the tradition is still not to include training in research and statistics methodology at the Bachelor's level in programs preparing students to become teachers. In many colleges and universities, courses in educational measurement and/or evaluation are included. Most of these courses, however, contain very little research or statistics methodology. Interestingly, each year more and more disciplines are requiring courses in both research methodology and statistics methodology.

At Winona State University, for example, the following departments or majors at the undergraduate level require both a research methodology and a statistics course: Communication Studies, Criminal Justice, Environmental Science, Exercise Science, Marketing, Nursing, Psychology, Social Work, and Sociology. A few others require only a research methodology course and several departments require only statistics.

For the period from 1987 to 2000, the number of sections (of approximately 37 students each) of introductory statistics in the Department of Mathematics and Statistics increased from 10 sections to 30 sections. As for teaching majors, only the major in Mathematics requires a statistics course and none of the teaching majors requires a research methodology course. It is important that all programs training teachers, but especially those training teachers in Special Education, require courses in research methodology and statistics as part of their degree programs so that those teaching have

preparation equivalent to those with non-teaching degrees and are better prepared for graduate level courses.

Interestingly, at Winona State University, starting in January 2000, an introductory course in statistics is being strongly recommended by the Department of Special Education for all its first-degree students preparing to be special education teachers. Further, an educational statistician from another institution contacted the author for suggestions for textbooks for their new Bachelor's level course in statistics in the College of Education since that College of Education would begin teaching statistics at the Bachelor's level for the first time starting in January 2001. Hopefully, these two additions of Bachelor's level statistics for education majors are not isolated cases, but indicative of a trend. It is too early to tell.

A final note

In this paper the author has discussed what she feels is the ideal set of topics and depth of coverage for research methodology and statistics courses for teachers. The course outline and various projects that the author uses have not been included because she is only allowed a total of 25 hours to teach the Master's students in special education both research methodology and statistics, of which 1 1/2 hours are spent having a librarian discuss how to use the library and the Internet to locate source material needed for scholarly reviews of the literature. Hence, her assignments, examinations, and projects are compromises between what she believes should be taught and what she can get done in 25 hours of classroom instruction with about 75 hours of extra work outside of the classroom expected of the students.

APPENDIX: PARTIAL LIST OF THESES AND PROFESSIONAL PAPERS UNDER THE MASTER'S DEGREE PROGRAM REQUIRING THE RESEARCH METHODS COURSES AT THE DEPARTMENT OF SPECIAL EDUCATION, WINONA STATE UNIVERSITY

THESES

Dennison, M. I. (2000). *Pre-service teachers self-reported perceptions of their computer skills computer with demonstrated skills*.

Griffin, K. A. (1997*). A validation of guidelines for the Special Education paraprofessional-teacher team and the student teacher*.

PROFESSIONAL PAPERS

Drazkowski, M. J. (1996). *An analysis of a secondary regular education initiative program.*

Foss-Zylstra, R. S. (1995). *Team teaching to meet the needs of students with the dual label of learning disabled and emotionally behaviourally disturbed*

Grace, J. (1993). *A survey of LD, ED, and CD teachers on their use and opinions of peer tutoring.*

Griffin, R. M. K. (1995). *Effects of repeated readings on the fluency, comprehension and attitude of learning disabled and non-learning disabled readers using a standard fifth grade Social Studies text.*

Gu, W. Y. (1997). *A comparison of Chinese and American fifth-grade samples on measures of arithmetic calculation*.

Hansen, H. (1995). *The effect of review cycles on retention.*
Hill, B. B. (1991). *Eye pursuit characteristics of second grade and fourth grade children.*
Holland, D. (1995). *Acquisition and retention of spelling words for students with mild disabilities using pattern generalisation and frequency word lists*
Jordan, J. (1995). *Descriptive study of educational recommendations given by a Midwestern university hospital clinic*.
Junker, J. (1999). *State self-esteem as effected by teacher mood states.*
Kreisel, C. (1993). *A follow-up survey of Special Education program completers at Winona State University from 1973 to 1984.*
Morris, A. N. (1998). *Assessing the effectiveness of an in-class delivery model for serving students with behaviour problems in the regular classroom: The intervention support team model of the Kenai Peninsula Borough School District Special Services Department.*
Navarre, D. (1990). *A comparison of teacher-student perceptions of the student's locus of control.*
Nyre, M. (1995). *The relationship between Attention Deficit Hyperactivity Disorder and learning disabilities.*
Peterson, C. (1993). *Teacher and parent perceptions of the effectiveness of mainstreaming exceptional educational needs students in a rural Midwestern school district.*
Radatz, D. (1991). *A survey of the post-school outcomes of learning disabled students.*
Traxler, K. (1999). *Family eating patterns and school performance.*

REFERENCES

Balboni, G., & Pedrabissi, L. (2000). Attitudes of Italian teachers and parents toward school inclusion of students with mental retardation: The role of experience. *Education and Training in Mental Retardation and Developmental Disabilities*, 35(2), 148-159.

Baumberger, J. P., & Bangert, A. W. (1996). Research designs and statistical techniques used in the *Journal of Learning Disabilities*, 1989-1993. *Journal of Learning Disabilities,* 29(3), 313-316.

Bruininks, R. H., Wolman, C., & Thurlow, M. L. (1990). Considerations in designing survey studies and follow-up systems for special education service programs. *Remedial and Special Education,* 11(2), 10-17 & 46.

Burrows, D. E., & Baillie, L. (1997). A strategy for teaching research to adult branch diploma students. *Nurse Education Today,* 17(1), 39-45.

Calder, J. E., Justen, J. E., & Waldrop, P. B. (1986). Research requirements in special education programs. *Teacher Education and Special Education*, 9(2), 51-54.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago: Rand McNally.

Cooke, N. L., Test, D. W., Heward, W. L., Spooner, F,, & Courson, F. H. (1993). Teachers' opinions of research and instructional analysis in the classroom. *Teacher Education and Special Education,* 16(4), 319-328.

Crowley, E. P. (1994-95). Using qualitative methods in special education research. *Exceptionality*, 3(2), 55-69.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal,* 6(3), 383-401.

Elmore, P. B., & Woehlke, P. L. (1998). Twenty years of research methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research. Paper presented at the *Annual Meeting of the American Educational Research Association* (ERIC Document Reproduction Service No. ED 420 701).

Estepa, A., & Sanchez Cobo, F. (2001). Empirical research on the understanding of association and implications for the training of researchers. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 35-51). Granada: International Association for Statistical Education and International Statistical Institute.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction* (6th ed.). New York: Longman.

Gersten, R., Lloyd, J. W., & Baker, S. with others (1998). *Designing high quality research in special education: Group experimental designs*. Reston, Virginia, USA: Council for Exceptional Children (ERIC Document Reproduction Service No. ED 420 957).

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. Thousand Oaks, California: Sage Publications.

Greenwood, C. R., Peterson, P. & Sideridis, G. (1994-95). Conceptual, methodological, and technological advances in classroom observational assessment. *Diagnostique,* 20(1-4), 73-99.

Harraway, J., Manly, B., Sutherland, H., & McRae, A. (2001). Meeting the statistical needs of researchers in the biological and health sciences. In C. Batanero (Ed.), *Training Researchers in the Use of Statistics* (pp. 177-195). Granada: International Association for Statistical Education and International Statistical Institute.

Heckenlaible-Gotto, M. J., & Choi, H. (1993). *The use of research methods and statistics in TECSE, 1990-1993: A content analysis. (*ERIC Document Reproduction Service No. ED 379 831).

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Hulek, A. (1983). Changing approaches to methodology of research in special education. *International Journal of Rehabilitation Research,* 6(1), 3-9.

Journal of Special Education. (1984), 18(1).

Journal of Special Education. (1985-86), 19(4).

Kratochwill, T. R., & Levin, J. R. (1992). *Single-case research design and analysis: New directions for psychology and education*. Mahwah, New Jersey: Lawrence Erlbaum.

Maxwell, S. E., & Delaney, H. D (1999). *Designing experiments and analysing data* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum.

McWilliam, R. A. (1991). *Mixed method research in special education*. (ERIC Document Reproduction Service No. ED 357 554).

Michael, E. (1989). Special education in Israel. *International Journal of Special Education,* 4(1), 59-64.

Stanovich, P. J., & Stanovich, K. E. (1997). Research into practice in special education. *Journal of Learning Disabilities,* 30(5), 477-481.

Swanson, H. L., & Alford, L. (1987). An analysis of the current status of special education research and journal outlets, *Remedial and Special Education,* 8(6), 8-18.

Tawney, J. W., & Gast, D. L. (1990). *Single-subject research in special education*. Paramus, New Jersey: Prentice Hall.

Todd, R. E., & Reece, C. C. (1990). *Graduate introductory educational research courses: A look at what is required.* (ERIC Document Reproduction Service No. ED 324 274).

Torgesen, J. K., & Dice, C. (1980). Characteristics of research on learning disabilities. *Journal of Learning Disabilities,* 13(9), 531-535.

Carol Joyce Blumberg
Department of Mathematics and Statistics
Winona State University
Winona MN 55987-5838, USA
E-mail: cblumberg@winona.edu