Acta Universitatis Wratislaviensis No 4242
Anglica Wratislaviensia 62.1, Wrocław 2024
https://doi.org/10.19195/0301-7966.62.1.10

Joanna Pitura
ORCID: 0000-0003-1659-1995
Institute of English Studies
University of the National Education Commission, Kraków, Poland
joanna.pitura@uken.krakow.pl

Teaching Quantitative Data Analysis with GNU PSPP: A Cognitive Apprenticeship Approach

Abstract: The ability to analyse quantitative data using software is of great importance but represents a substantial challenge for many in the applied linguists' community. One way to support them in this challenge is to provide inexperienced analysts with appropriate training. Accordingly, this methods tutorial paper aims to present the GNU PSPP open-source software and provide guidance on the development of analytic skills using this software among higher education students. The cognitive apprenticeship theory and research are addressed first as the theoretical foundation. What follows is a description of how cognitive apprenticeship approach was applied to teaching analytic practice in a Moodle-based course offered to graduate students enrolled in the TEFL MA programme at a university in Poland, where GNU PSPP was integrated in the coursework. The paper concludes with a brief evaluation of the course, showing that this approach works well for supporting students throughout the course and for enabling them to transition towards conducting quantitative data analysis using software in their own MA research projects. The paper may be of interest to those tasked with providing training and development programmes to students, faculty and researchers in applied linguistics and beyond, including those for whom access to commercial software may be of concern.

Keywords: quantitative data analysis, GNU PSPP, cognitive apprenticeship

1. Introduction

Quantitative data analysis—i.e., the analysis of data represented by numbers (Creswell 175; Dörnyei 199; Lowie and Seton 13)—is one of the stages in the research process, preceded by the formulation of a research problem, the planning of research, and data collection, and followed by reporting and evaluating research (Creswell 174–201; Dörnyei; Wild and Pfannkuch 226). The data analysis stage itself is conducted in a series of steps, from preparing the data for analysis (e.g., assigning numeric values to the data and inputting the data into data analysis software), through running the analysis (e.g., conducting descriptive and inferential statistical procedures), reporting the results (using tables and figures), to discussing the results obtained in the analyses (Creswell; Dörnyei). Software packages—paid (e.g., SPSS, SAS, STATA) or free, open-source (e.g., GNU PSPP)—aid researchers in their analytical process. These allow researchers to conduct a wide range of statistical procedures, including descriptive and inferential statistics, as well as more advanced procedures, such as regression or factor analysis, also allowing for data visualization through tables, charts, and histograms. The ability to analyse quantitative data with the use of software is essential for quantitative and mixed methods researchers (cf. Loewen et al., "Knowledgeable" 886), and may also be of great importance for qualitative researchers conducting, e.g., case or action research studies, who may need to analyse quantitative data (e.g., participants' test scores) (Pitura, *Using*). Without the ability to independently analyse, interpret and present quantitative data, researchers might be unable to deliver trustworthy results that will contribute to the body of knowledge in the field, enhance practice, and guide policy decisions (Creswell; Gass et al. 252; Loewen et al., "Knowledgeable" 887).

Regrettably, existing research suggests that many researchers in the field of applied linguistics may have insufficient knowledge and skills for conducting such analyses effectively and appropriately on their own. Research done on researchers, faculty, and graduate and doctoral students in the area of applied linguistics revealed deficiencies in their statistical literacy, despite having attended courses in statistics, which, *nota bene*, the majority of the surveyed participants found not adequate and satisfactory (Gönülal; Gönülal, Loewen, and Plonsky; Lazaraton et al.; Loewen et al., "Knowledgeable", "Statistical"). This situation exists in spite of the availability of many handbooks, textbooks, chapters (e.g., Dörnyei; Larson-Hall, *A Guide*; Lowie and Seton; Mackey and Gass; McKinley and Rose; Phakiti et al.; Plonsky), and articles (Larson-Hall, *Moving*; Larson-Hall and Herrington; Lindstromberg; Norris; Norris et al.; Plonsky and Oswald) written, among others, to improve applied researchers' quantitative data analysis skills. This brings to the foreground the significance of suitable training and development programmes for students, faculty, and researchers allowing for the development of quantitative data analysis skills that they need to understand and assess the quality of research published in the field and to conduct their own research projects (Gass et al. 254; Loewen et al., "Knowledgeable" 887).

Importantly, it has been pointed out that data analysis is, at least in part, a craft skill that is best acquired by hands-on experience (Li and Seale). From this perspective, the theory of cognitive apprenticeship (Collins et al., *Teaching Students*; Collins et al., *Making Thinking*; Collins and Kapur), which is gaining traction in educational contexts (Matsuo and Tsukube), may be a useful instructional approach to teach quantitative data analysis skills. It emphasizes the importance of interaction with a competent practitioner who models the target cognitive activity for learners to observe, guides learners while they practise, and provides them with feedback on their performance. In order to improve graduate students' problem-solving in educational statistics, this theory was applied to design a Moodle-based instructional intervention, which appeared to have beneficial effects on student learning of statistics (Saadati et al.). However, while this unique research shows that the theory of cognitive apprenticeship can be considered an adequate teaching approach to help students understand statistical concepts and procedures, there is not enough information available on how to integrate quantitative data analysis software in the coursework to more fully develop students' capacity for data analysis. Without such guidance, research methods educators might not be able to assist students, faculty, and researchers in gaining expertise in quantitative data analysis in university courses and professional development programmes.

In efforts to address this challenge, the purpose of this methods tutorial paper is to promote the teaching of quantitative data analysis with the use of software in the field of applied linguistics. In light of this, the cognitive apprenticeship theory and research are presented first. What follows is a description of how cognitive apprenticeship approach was applied to teaching analytic practice in a Moodle-based course offered to graduate students enrolled in the Teaching English as a Foreign Language (TEFL) MA programme at a university in Poland. The course utilized the GNU PSPP open-source software due to limited access to commercial software in this context. In doing so, this work adds to the body of methodological literature available, and responds to calls for improving methodological skills in the field of applied linguistics (Gass et al. 2021; Loewen et al., "Knowledgeable").

2. Cognitive apprenticeship—theory and research in education

The cognitive apprenticeship model was developed by Collins and colleagues (Collins et al., *Teaching Students*; Collins et al., *Making Thinking*; Collins and Kapur) as a variation on the traditional apprenticeship that focuses on workplace training in the field of crafts and trades (e.g., shoemaking). In this context, the targeted skills are easily observable from the outside, allowing apprentices to learn the skill by watching the master perform a specific task and practise with master's support. Instead of focusing on physical skills, the model of cognitive apprenticeship concentrates on cognitive skills, such as the learning of mathematics, reading, or writing. The model is based on the premise that, in contemporary classrooms, students do not have access to teacher's mental processes, which can serve as a basis for learning through observation. Similar to this, teachers are unable to observe students' thought processes, making it impossible to monitor how students are using knowledge to complete the tasks at hand. Therefore, the aim of cognitive apprenticeship is to make the thought processes used by experts (teachers, experienced practitioners, professionals) while performing cognitive tasks visible so that students can see them in action and practise them. It is maintained that knowledge is valuable when it can be employed in various situations to solve problems in the real world. As such, cognitive apprenticeship theory appears to be suitable to teach quantitative data analysis skills, allowing students to acquire knowledge and skills under the guidance of instructors before applying these skills for their own research project purposes.

The theory distinguishes four elements constituting the cognitive apprenticeship learning environment: Content, Method, Sequence, and Sociology. Each element is described in more detail below.

The first element, Content, covers the many kinds of knowledge novices need for expert competence. This area encompasses *domain knowledge*, i.e., concepts and procedures specific to a field of study, and *strategic knowledge*, i.e., knowledge of how to apply domain knowledge to address real-world problems. Strategic knowledge, in turn, comprises a range of strategies, such as *heuristic strategies* (techniques for completing tasks), *control strategies* or *meta-cognitive strategies* (approaches for regulating task execution), and *learning strategies* (knowing how to learn concepts and procedures) (Collins and Kapur 111–12). It is thus clear that, in order to effectively teach students how to analyse, interpret, and present quantitative data in their own research, instructional content should place a strong emphasis on helping them become familiar with statistical concepts and procedures, conducted with the aid of statistical software.

Method, the second element in the cognitive apprenticeship framework, highlights the importance of providing students with the chance to observe and learn from experts (teachers) in context. In light of this, the model offers six distinct methods for fostering the growth of expert knowledge, i.e., modelling—students observe the teacher completing a task; coaching—student work is observed and facilitated by the teacher; scaffolding—students are assisted by the teacher in completing a task; articulation—students are encouraged by the teacher to express their understanding verbally; reflection—students compare their own performance with others' performance on a task; and exploration—students formulate and address their own problems (Collins and Kapur 113–15). It is assumed that the first three methods—modelling, coaching, and scaffolding—aim to help students in acquiring skills through observation and practice under expert's (teacher's) guidance, which is similar to what takes place in traditional apprenticeship. The methods of articulation and reflection are provided to assist students in gaining awareness of and control over problem-solving processes, whereas the method of exploration aims to promote autonomy by allowing students to formulate problems on their own. These methods seem to have the potential to support the learners of quantitative data analysis by giving them the chance to interact with and learn from experienced researchers in authentic research contexts.

The third element, Sequencing, addresses the issue of organizing instructional activities. The theory holds that tasks should be of increasing complexity, increasing diversity, and ought to focus on global before local skills. Specifically, it is maintained that students need to: 1) be provided with learning activities that become progressively complex, 2) practice in a range of contexts, and 3) understand the entire task before carrying out its specific parts (Collins and Kapur 115). It follows that the objectives in the course that teaches quantitative data analysis should similarly focus on global knowledge before specific one, grade the introduction of statistical concepts and procedures, and should also employ a wider range of statistical concepts and procedures in learning tasks over time.

Finally, Sociology, refers to the social aspects of the learning environment and includes four characteristics. First, it makes reference to situated learning, wherein students gain knowledge while working on practical problems. Next, Sociology embraces the notion of community of practice which entails the discussion of the different ways tasks can be performed. Intrinsic motivation, in turn, concerns students creating their own objectives to learn skills. The last aspect, exploiting cooperation, emphasizes student collaboration while achieving their objectives (Collins and Kapur 115–16). Clearly, such social factors must also underpin the design of the quantitative data analysis course.

The principles of cognitive apprenticeship have been applied in the design of learning environments in a range of educational contexts, as reported in a recent review study conducted by Matsuo and Tsukube. The researchers found that, in primary and secondary education settings, the theory has been used for designing and implementing programmes in science, chemistry, and multimedia education. The theory has also been used in post-secondary education to create teacher-student tuition programmes for music instruction, performance consulting programmes in management education, qualitative research methods in programmes for doctoral students, and grant proposal writing programmes. In addition to this, there are programmes for pre-service and in-service teachers, as well as those in healthcare, that are based on cognitive apprenticeship theory. The theory has been employed in adult education for a variety of purposes, including the creation of writing curricula for adolescents, military education programmes, business education programmes for the hospitality and tourism industry, and library training. Finally, as reported by the researchers, the theory has been applied to develop Internet-based programmes, such as a writing programme in primary education, an instructional planning programme in pre-service teacher education, and a statistics course in postgraduate education. Overall, existing evidence presented in this review study demonstrates the value of cognitive apprenticeship theory in programme design, web-based programmes in particular, for enhancing the acquisition of cognitive skills among students and employees. As the study describing the design of the statistics course is of special relevance to the current paper, it will be discussed in more detail next.

In efforts to improve the learning of educational statistics among postgraduate students, Saadati et al. conducted a study to evaluate the effectiveness of a Moodle-based learning environment designed as an extension to face-to-face lectures, referred to by the researchers as a web-based Cognitive Apprenticeship Model (or *i*-CAM). In the main, the designed model of teaching consists of three phases: handling, supporting, and self-exploring. At the handling phase reported in the paper, congruent with the methods of modelling and coaching of the cognitive apprenticeship theory, instructors use intentionally arranged learning resources and activities available on Moodle to demonstrate statistical procedures, students complete problem-solving tasks, and leave comments in the forum if they have questions. At the supporting phase, consistent with the methods of scaffolding, articulation, and reflection, scenarios describing statistics problems are posted in the forum. Students follow instructors' hints (that decrease in amount over time) to arrive at solutions. In addition, students explain their thinking about how to approach a given statistical problem and point out the challenges they may be facing. In addition, students engage in conversations with one another to reflect on the work they complete. In the final, self-exploring phase, students are encouraged to talk about the difficulties they had with their assignments and share their solutions. The study participants comprised two groups of postgraduate students from the Faculty of Educational Studies enrolled in the Educational Statistics course: the treatment group ($n = 27$) and the control group ($n = 26$). Both groups had the same content, assignments, instructor, and class duration. What differed was the way students conducted the learning activities on Moodle; while the treatment group was guided by instructors in the learning activities in the way described above, the students in the control group had access to the materials on Moodle, engaged in discussions with peers, and could also approach their instructors for help with their tasks. Data were collected with a pre-test and post-tests to measure students' knowledge of statistics at the end of every phase of the *i*-CAM. In terms of mean test scores at the end of each phase, the treatment group outperformed the control group, indicating that the programme had enhanced statistical problem-solving skills of the participants in the treatment group.

All in all, the theory of cognitive apprenticeship seems to help structure our thinking about how to design and implement the training of cognitive skills, quantitative data analysis training programmes in applied linguistics being a case in point. Importantly, this theory contributes to our understanding of quantitative data analysis abilities as a complex cognitive skill, requiring the knowledge of statistical concepts and procedures that can be next applied to conduct statistical procedures with software, to interpret and present the obtained results. The theory also offers helpful information regarding the kind of learning environment that is conducive to the development of analytic skills. The study describing the effectiveness of the web-based statistics programme offers particularly valuable insights on how the principles of cognitive apprenticeship can be applied to design a learning environment aiming at the development of analytic sills. However, it appears that research systematically and purposefully integrating software to develop students' ability to conduct statistical procedures has not been conducted to date. Without such research, instructors might be left with little guidance on how to create training programmes for quantitative data analysis that are effective. As a result, without exposure to the use of quantitative data analysis software, students, faculty, and researchers might find it more difficult to become independent analysts. This paper extends the existing literature by embodying the principles of cognitive apprenticeship theory in the design of a Moodle-based course integrating GNU PSPP in the coursework. In what follows, there is a description of the course introducing MA TEFL students to qualitative data analysis with open-source software, aiming to assist students in understanding how the instructor, an experienced practitioner, creates analytical workflows with the use of the software, and to help students gain the ability to apply analytic skills in their own MA research projects.

3. Moodle-based quantitative data analysis instruction for MA TEFL trainee teachers in Poland

3.1. Background

The Institute of English Studies at a university in Poland offers the IT in Didactic Research course to MA TEFL students who are required to write and defend a research-based MA thesis. This course introduces students to both quantitative and qualitative data analysis with software in TEFL research. The said course is a continuation of the Research Methods in Language Education course taught a year earlier which aims to familiarize the students with the issues around various research approaches, designs, and data collection methods, covering, among others, research ethics, forming research questions, and so forth. The number of participants in the course fluctuates each year, the maximum being twenty students in a group. The course has been taught by the present author since 2016 in the face-to-face mode, except for the 2020–2021 and 2021–2022 academic years when instruction was held synchronously online via MS Teams. Course content and delivery has evolved over years to accommodate students' learning experiences, as well as author's own experiences as a course instructor and a supervisor of MA theses. In what follows, the focus is on the first part of the course in its current form that deals with quantitative data analysis utilizing the GNU PSPP open-source software instead of a commercial statistical package, since students lack access to the latter.

3.2. Course goals

Course goals target the development of domain knowledge and procedures necessary in the process of quantitative data analysis. Specifically, in line with the regulation of the Polish Minister of Science and Higher Education on the standard of education preparing for the teaching profession (*W sprawie standardu*), course goals aim at developing students' expertise with regard to:

- principles of quantitative data processing and analysis,
- verification and coding of quantitative data,
- basic statistical analysis, e.g., descriptive statistics, frequency distributions, measures of central tendency and dispersion, correlations, statistical inference,
- the use of software supporting the analysis of quantitative data.

3.3. Theoretical framework

Course design is informed by the cognitive apprenticeship theory (Collins et al., *Teaching Students*; Collins et al., *Making Thinking*; Collins and Kapur) which involves acknowledging the cognitive character of the abilities underpinning quantitative (and qualitative) data analysis. Therefore, it is recognized that the learning environment must allow students to have experiences akin to those of an apprentice, to observe the instructor working, and to practise under instructor's guidance and apply statistical concepts and procedures using software in their own research. Additionally, it is considered essential that, in order for students to advance toward independent skilful performance, they need to be surrounded by both experienced researchers-analysists and other students, who are all engaged in data analysis in real-life research that constitutes one of the stages of the research process (Creswell; Dörnyei; Wild and Pfannkuch).

3.4. Course resources

The course is supported by a number of online resources available on/through Moodle—the university Learning Management System. These include:

- GNU PSPP software download links and installation instructions,
- selected chapters from the electronic version of the statistical handbook (in Polish) edited by Bedyńska and Cypryańska, accessible online through the university library,
- e-articles reporting on empirical research in the field of TEFL,
- instructor's Google Slides class presentations, mainly based on the content from Bedyńska and Cypryańska,
- YouTube video tutorials demonstrating specific statistical procedures with the use of GNU PSPP recorded by the instructor,
- sav. file datasets for student practice,
- GNU PSPP tasks prepared by the instructor requiring students to conduct analyses with the use of GNU PSPP, the provided dataset, and a video demonstrating how the analytic task is to be accomplished, as well as

specifying how the output of analyses is to be made available for instructor's feedback.

In addition, in order to keep student work organized and accessible for the instructor throughout the course, Google Docs are used to serve as student online workbooks. Students create their own documents and share the link on the Workbooks Moodle forum.

3.5. Content

The content is sequenced and designed to be covered in separate classes as Moodle modules in the following way:

- Class (module) 1. Preparing data for analysis in PSPP,
- Class (module) 2. Data presentation—Frequencies,
- Class (module) 3. Data presentation—Descriptives,
- Class (module) 4. Scales and indicators,
- Class (module) 5. Correlation,
- Class (module) 6. Statistical significance,
- Class (module) 7. Differences—Categorical variables,
- Class (module) 8. Differences—Continuous variables.

It is assumed that the mastery of these concepts and procedures is necessary for novice researchers on their way to expert performance.

In each class (module), learning resources are organized to support the teaching of specific statistics concepts and procedures, in keeping with the tenets of cognitive apprenticeship. For example, in a class focusing on correlation, the resources available on Moodle include an e-article reporting on a research study utilizing correlational analyses (resource 1), a Google Slides presentation on correlations (resource 2), tasks in which students learn how to conduct correlational analyses with PSPP (resources 4 and 5) using the provided dataset (resource 3), as well as instructions on reading materials to prepare for the forthcoming class (resource 6), i.e., a chapter from the statistical handbook and a new research e-article (Figure 1).

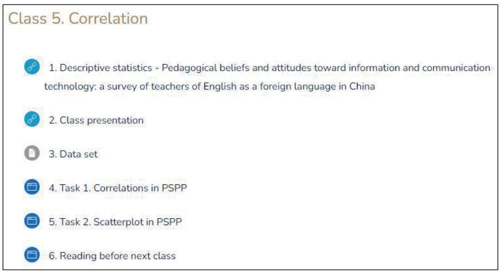

Figure 1: Organization of the resources in a class (module) devoted to correlational analyses with GNU PSPP

3.6. Procedure

Class procedure is designed to support the development of students' analytic skills using the six methods derived from the theory of cognitive apprenticeship in four phases, as presented in detail below. In the main, this course differs from the course described by Saadati et al. (2015) in that all the six methods of cognitive apprenticeship are applied to promote the learning of each concept or procedure in every class, while the said researchers applied the theory to structure the learning activities in three phases throughout the whole course.

3.6.1. Phase 1: Building students' knowledge of concepts and procedures in quantitative data analysis

Phase 1 focuses on developing students' knowledge regarding a specific statistical concept or procedure necessary for students to be able to conduct quantitative data analyses. During this phase, the activities are organized to correspond with the methods of modelling, scaffolding, and coaching of cognitive apprenticeship.

Specifically, before class, in order to prepare for learning in class, students are asked to read two texts: 1) a chapter form a handbook of statistics (Bedyńska and Cypryańska 2013); and 2) a research article that utilizes that specific statistical concept or procedure. For example, when preparing for class dedicated to correlations, students read a chapter entitled "Measures of relationship between variables—correlation coefficient" (Bedyńska and Cypryańska) and an article "Pedagogical beliefs and attitudes toward information and communication technology: a survey of teachers of English as a foreign language in China" (Liu et al.).

In class, the instructor begins the learning sequence by analysing the article. He/She asks questions such as: What is the purpose of the study?, Who were study participants?, What data collection methods were employed?, What were the findings?, etc. Accordingly, in a class aiming to familiarize the students with correlations, students answer these questions about the study on teachers' pedagogical beliefs and attitudes towards ICT. By referring to the whole research inquiry cycle conducted by experienced researchers, the instructor contextualizes the use of a specific analytic practice in a wider research process and, in so doing, prevents student perceptions of a target statistical concept or procedure as not a particularly useful theoretical construct.

Next, the instructor focuses students' attention on the Results section in the article that features a target statistical concept or procedure. The instructor asks: Do you recognize the statistical procedure used here?; Why do you think it was used?; etc. These questions are asked to explain the need for this type of analyses with reference to the research questions formulated by the researchers—authors of the analysed article. For example, in the article utilizing correlational analyses, the instructor draws students' attention to the correlational matrix and the interpretation found in the text and elicits the reason why this type of analysis was conducted.

As such, the research article helps model data analysis by demonstrating how experienced researchers conduct their analyses in real-life research context in order to establish the need for student ability to conduct specific analyses and to motivate students to make effort to engage in the activities that follow.

After that, the instructor organizes and develops student knowledge concerning the target concept or procedure using a Google Slides presentation. Throughout the presentation, the instructor guides the students to constructing the understanding of specific statistical concepts or procedures. Specifically, he/she explains the concept or procedure, applies it in examples, and clarifies difficulties. For example, to develop student understanding of correlation, the instructor gives a presentation covering the following aspects:

- introducing students to correlation,
- cautioning students that correlation does not imply causation,
- providing basic characteristics of *r*-Pearson correlation coefficient,
- explaining how specific values are interpreted,
- displaying examples of scatterplots as graphic representations of correlation,
- listing other correlation coefficients (rho-Spearman, Kendall’s Tau-b, Phi, Cramer’s *V*, and Eta),
- providing students with an example of a correlation matrix and of a report.

On the whole, the procedures in this phase allow for building student domain knowledge that underpins the ability to analyse quantitative data, interpret the output and present the results with regard to a specific statistical concept or procedure. This phase resembles the handling phase described by Saadati et al. in that students are engaged in the activities that aim at developing their understanding of statistical problem solving using the learning resources and activities available on/through Moodle. However, in the course described here, this phase constitutes an initial part of each class—not the whole course—providing students with the theory and practice of each targeted statistical concept and procedure.

3.6.2. Phase 2: Improving students’ ability to analyse quantitative data with GNU PSPP

Phase 2 aims at developing students’ knowledge concerning the use of GNU PSPP to conduct a specific statistical procedure. The purpose of instruction is to help students deepen their knowledge of data analysis and develop their analytic skills through the observation of the instructor performing analyses with the use of software and the practice under instructor’s guidance. Like the previous phase, phase 2 is aligned with the methods of modelling, scaffolding, and coaching to improve students’ skills in the use of quantitative data analysis software.

In class, the instructor makes use of video-based GNU PSPP tasks available on Moodle to meet this objective. Individually, using headphones, students complete GNU PSPP tasks, guided by instructor demonstration recorded in videos.

They take screenshots of the output and paste them in their online workbooks. For example, when covering the topic of correlations, students are asked to conduct correlation analyses with GNU PSPP on the dataset provided by the instructor by completing two tasks. Task 1 involves correlating variables (Fig. 2), in which the output—the corelation matrix—is pasted in the workbook. Task 2, in turn, entails generating a scatterplot for two variables and pasting the output in the workbook as well (Figure 3). The students who are absent, need more time to understand and perform a specific procedure, or, for technical reasons, are unable to complete tasks in class, are allowed to complete and submit tasks after class.

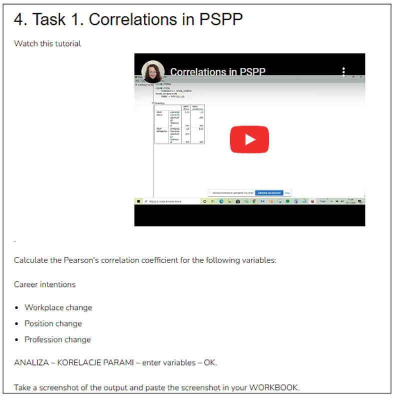

Figure 2: Conducting correlations in PSPP—Task 1 instructions

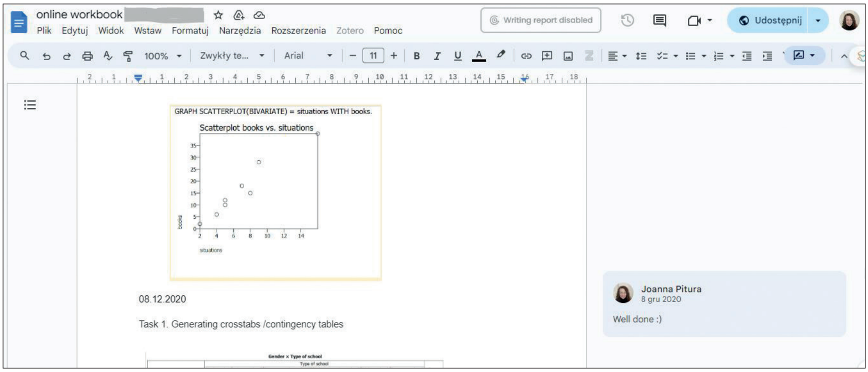

Figure 3: Instructor feedback provided as a comment in a student workbook

This procedure allows the instructor to support each student individually in observing an experienced analyst perform a target procedure utilizing GNU PSPP and to scaffold student emerging analytic performance. In particular, video tutorials help the instructor to verbalize the procedures applied while completing the task. In addition, by making the demonstration videos available online, students can access the video tutorials whenever necessary and replay them as many times as they need. In this way, support is individually adjusted and becomes less and less necessary over time.

As students work on PSPP tasks, the instructor monitors progress on students' computer screens and checks the output provided in student online workbooks. The instructor assists those students who need help, provides feedback face-to-face and through comments in student workbooks. An example of a comment provided in a class covering correlation is displayed in Figure 3. In this way, the instructor promotes competent analytic practices among the students.

Unlike Saadati et al.'s course, this course includes content allowing students to acquire the ability to conduct quantitative data analysis with the use of software. It is believed that without this content, students' domain knowledge would be incomplete.

3.6.3. Phase 3: Increasing students' control of analytic practice

Phase 3 concentrates on increasing student control of analytic practice with regard to a statistical concept or procedure covered in class. At this phase, corresponding with the methods of articulation and reflection of the cognitive apprenticeship theory, learning activities assist students in developing the ability to interpret obtained results.

This phase begins when students have completed GNU PSPP tasks and submitted the output in their workbooks. The instructor displays the output then, students compare their results with the one displayed, and are asked to interpret this output. If necessary, the instructor modifies student interpretation, ultimately refining their understanding of a specific statistical concept or procedure. For example, when students complete correlation tasks, the output (displayed in Figure 3) is interpreted (e.g., The more books students read, the more situations they can deal with). After that, the instructor goes back to the article by Liu et al. analysed at the beginning of this class to interpret the results obtained by the researchers. Students look at the correlation matrix shown in the article and comment on the strength and direction of a few selected results. By giving students opportunities to express their understanding verbally, i.e., interpreting the results of the work they have completed and the results reported in the article, as well as by allowing students to compare their interpretation with experienced researchers' interpretation, the instructor helps students gain increasing control over their own analytic practice.

This phase echoes the Supporting Phase described in Saadati et al.'s study, during which students were asked to work with other students, to compare their knowledge with that of others, to report their thinking while solving statistical problems and the challenges they faced while performing. In the present course, the cognitive apprenticeship methods of articulation and reflection are applied in each class to develop students' ability to use each statistical concept and procedure, which is different from Saadati et al.'s course, in which these methods were applied in the second phase of the whole course, covering a range of statistical concepts and procedures.

3.6.4. Phase 4: Promoting student independence in analytic practice

Phase 4 aims to foster student use of a specific statistical concept or procedure in a different context, i.e., the content of the research goals pursued in their own research projects. This phase aligns with the method of exploration and is designed to give students opportunities to (re)formulate problems requiring the application of quantitative data analysis skills, and to conduct such analyses on their own.

At this phase, with instructor's support, students focus on their own data collection and analysis plans and explore the applicability of the concept or procedure taught in class in their own research. They are encouraged to revise their plans, having gained more knowledge about a specific statistical concept or procedure. For example, in a correlations class, students are asked to list all (planned) continuous variables in their study, sketch a hypothetical correlation matrix, and fill it in with fictional output. They share their visualizations with peers and the instructor, who provides additional guidance and advice, if needed. As such, students explore the pertinence of a specific concept or procedure in the context of their own research in order to continue developing their competence in quantitative data analysis.

This phase resembles the last phase in the course designed by Saadati et al., during which students continued their learning activity while cooperating with peers, the instructor remaining withdrawn and providing far less assistance. Again, in this course, the exploration of how a particular concept or procedure may be used in different contexts is conducted in every class allowing students to investigate the applicability of target concepts or procedures in the context of their own research plans.

3.7. Assessment

Students' analytic skills are assessed in the written exam conducted on Moodle in a university computer lab at the end of the course. The exam consists of two parts: theoretical and practical. The first part of the exam covers the theory of quantitative (and qualitative) data analysis. Students are provided with true/false items, such as "The Pearson correlation coefficient has a value between 0 and +/−1 inclusive" (true / false) or "T-test are used to compare three or more means" (true / false). The practical part requires the students to conduct analyses with the use of GNU PSPP and report their results in the exam form. Sample task instructions are displayed in Figure 4, and examples of questions testing students' ability to conduct statistical procedures with software are presented in Figure 5.

PSPP TASK

A researcher studied English language teaching effects in two schools, taking three skills into consideration: 1) Speaking, 2) Writing, 3) Reading. A group of 10 students took 3 tests to measure each skill, the rating scale being 0–50 for each test. Find the results below.

Student	School	Speaking	Writing	Reading
1	1	34	24	24
2	1	31	19	50
3	1	25	24	29
4	1	47	10	20
5	1	29	25	38
6	2	20	21	45
7	2	50	27	23
8	2	34	29	14
9	**2**	15	17	29
10	2	43	29	46

Step 1. Create the following variables in PSPP: School, Speaking, Writing, Reading.

Step 2. Enter the data for 10 participants.

Step 3. Answer the questions included in the test.

Figure 4: Sample GNU PSPP examination task

What is the value of the standard deviation (*SD*) of the variable Writing for students in School 1?

a. 6.27
b. 5.37
c. 6.48
d. none of the above

What is the value of the Pearson's *r* coefficient for Speaking and Writing?

a. 0.129
b. −0.29
c. 0.456
d. none of the above

Figure 5: Sample GNU PSPP examination questions

4. Evaluating course suitability—student perceptions

In order to evaluate the suitability of the course based on the cognitive apprenticeship model, this methods tutorial uses student stories that were submitted as part of their coursework. As the course progressed to the second part (which focused on qualitative data analysis), the students were asked to document their beliefs, feelings, and experiences related to statistical analyses before, during, and after their learning in the course. The intent of this assignment was to create a textual dataset that could be used to train students in qualitative data analysis. Out of the total number of students participating in the course ($n = 33$) in 2020–2021, twenty-five students (female $n = 22$, male $n = 3$) provided written informed consent for the use of this work at the conclusion of the course. For the present tutorial purposes, the stories were analysed to provide insights into students' perceptions of their competence, beliefs, and feelings around quantitative data (statistical) analysis using GNU PSPP (see also Pitura *Revealing* for a Systemic Functional Linguistics analysis of MA TEFL students' attitudes towards statistics in this course). In what follows, student words are quoted in the original form.

Based on the responses, before the course, the students' perceptions of statistics were shaped by their lack of familiarity with and experience in conducting research and quantitative data analysis. The students had limited knowledge of statistics (e.g., "I only knew what 'mean', 'mode', and 'median' were.") and had little to no experience using statistical tools beyond Microsoft Excel ("I know that thanks to this program we can create various diagrams and tables."). Yet, some recognized the importance of statistics in their MA thesis research work and felt a need to learn more about it. Some students expressed uncertainty about conducting research and analysing data, as well as a lack of confidence in their ability to understand statistical concepts (e.g., "I must confess that when reading research for my MA I would only glance at tables with data because I had no idea how to interpret what they presented."). Many students initially experienced negative feelings: they reported feeling intimidated, stressed, overwhelmed, terrified, and "not … too excited" by the prospect of taking a course teaching statistical concepts. Some of them were anxious about the expectations of the course and their ability to succeed in it (e.g., "I would say that I am rather a humanistic soul, that is why I was not so optimistic about these classes"; "I was perplexed and scared how would I ever succeed in such a course."). Some students reported feeling that statistics was a new domain and not something they expected to learn during their studies. Others expressed a dislike for numbers and maths (e.g., "a word 'statistics' sounded boring to me, so I did not think that I would enjoy these classes at all."). Only three students were positive about the prospect of taking a course on statistics. While one of them expressed curiosity and a desire to learn about the subject, another one expressed gratitude for the opportunity to learn statistics as they recognized its relevance to their future research work. One of the students expected that the course would be challenging, but they also viewed it as useful.

The data show that the first, introductory class was rather daunting for the students. For many, the concepts were new; one student reported having the impression of taking a lesson in a new foreign language and another student was unable to connect the content supplied in class with their research process. Some students expressed feelings of dislike (e.g., "During the first class in which we started to learn statistics I felt that it is not my cup of tea."), fear, and anxiety, with one even bursting into tears. However, despite these initial challenges, students' perceptions of statistics became more favourable over time. With the instructor's explanations, statistics became more understandable and tasks were more feasible. Participants reported learning how to enter, compare, and visualize data, as well as gaining familiarity with statistical vocabulary in English. Although some struggled while learning (e.g., "Every class I had a feeling that I totally do not get what the teacher tried to convey. If I said that I understood everything, it would be a lie."), others found it ground-breaking to discover new software and websites ("As time passed by, I got familiarized with certain websites, [software] and discover what my laptop is capable of. It was ground-breaking."; "I had no idea that there was a special program to solve all my doubts as well as uncertainties—PSPP."). In addition, the participants considered the course to be practical and useful, with direct applicability to their MA thesis research. The data suggest that students' feelings shifted from initial overwhelm to a positive and enthusiastic outlook as they became more confident and interested in statistical concepts. As reported by one of the students, although they found the concepts complicated and had to spend extra time after class to study them, they ultimately derived satisfaction from grasping the ideas, one student even found enjoyment in creating charts and tables. Beyond this, some participants declared gaining respect for those working in research and recognized the importance of understanding statistical procedures for conducting research. One participant expressed a newfound "fascination towards this discipline and its power to organize what was disorganized and to label the terms that seemed impossible to be defined".

Having completed the course on quantitative data analysis, student words indicate that the course was challenging, but valuable and rewarding, providing them with valuable skills and knowledge which will be useful not only in their thesis writing, but also in future research. The students declared gaining theoretical and practical knowledge about statistics (e.g., "now, numbers do not scare me anymore, I know where to look, what they mean, how to read them"). In addition, the students reportedly developed confidence in their ability to understand research (e.g., "I do not regard statistics as something undoable"; "right now I do not find statistics intimidating"; "interpreting research papers would be way less daunting to me"), as well as to conduct their own research and analyse data (e.g., "I feel that I made a lot of progress, and believe that without the course I would have problems with hitting the deadlines I agreed on with my MA supervisor.”). Although some students still need to revise their knowledge (e.g., “As for the theory, I really need to go through my notes a few times more, since there are still some issues that I have not fully understood yet.”), they generally feel prepared to complete their MA thesis (e.g., “I feel that I have the knowledge to conduct my research and show its results in the right and correct way.”; “I feel that I have all the tools and a considerably big amount of knowledge needed to complete the statistical part of my work.”).

As to the tasks requiring the use of PSPP, it appears that the students had positive experiences, despite some initial difficulties. For example, one of the students reported that during the first lesson, they had problems with their computer, and they were very afraid that they would be unable to meet the deadline. Another student also indicated trouble completing the tasks; they got lost multiple times, in spite of being provided with video tutorials. It took longer than they had anticipated to complete the assignments. However, the students realized that they were given plenty of time to complete tasks and that the tasks themselves were manageable. Student words also indicate that the PSPP software was first met with some scepticism, (“I thought PSPP would be another non-user friendly program I would download, install, and never use again”). Overall, it appears that, with practice, the students became comfortable and skilled to use the software, the students found PSPP to be intuitive and a helpful tool for data analysis, and many planned to use it in future research projects.

Video tutorials were well-received by the students. In general, the tutorials were seen as a valuable resource which helped the students to increase the ability to use the software (e.g., “In fact, I found PSPP to be quite easy to figure out with the help of tutorials”; “thanks to the tutorials that the teacher prepared, those classes were not hard”). The tutorials were considered more helpful than live lessons as they provided the ability to fast-forward or rewind the video, allowing students to work at their own pace (“I must admit that such tutorials are much better than live lessons as I have the chance to fast-forward the video or go back if needed.”). The tutorials were also praised for their clarity and simplicity (“I also liked the fact that we had tutorials prepared specifically [by] the teacher that were really concise and easy to follow.”), and one of the students appreciated the instructor’s voice in the videos (“Additionally, I loved the voice of my teacher—it eased my mind every single time.”). The tutorials were especially helpful for those students who missed a class, as well as those who needed to review a concept.

Finally, individual students shared their experiences with other course resources, i.e., the handbook of statistics, Google Slides presentations, e-articles, the dataset, and the Google Docs workbook. As reported by one of the students, the first chapter selected from the handbook of statistics proved to be challenging, so the student sought help from their mother, a maths teacher, to be able to grasp the basics. According to another student, the presentations were helpful in clarifying theoretical concepts, but they had to revisit them multiple times to fully understand the presented aspects. E-articles, in turn, were considered valuable for one student's future MA thesis. The dataset provided by the instructor was described as "interesting" by yet another student as they "could see the differences between age and gender, schools, degrees among teachers". Lastly, one student found the workbook to be a convenient tool to refer back to while conducting research ("We have everything in one place and we can always go back there while doing our research to recall some information.").

5. Conclusion

Recognizing the importance of the ability to conduct quantitative data analysis with the use of software as a core research skill, this methods tutorial addresses the issue of teaching quantitative data analysis to students, faculty and researchers in the field of applied linguistics. In the effort to enhance methodological training, the paper presents a Moodle-based course that integrates software training in the coursework to help MA TEFL students acquire quantitative data analysis skills. The theory of cognitive apprenticeship was applied to design and deliver learning experiences allowing students to observe, imitate, and apply analytic procedures in their own research.

Based on students' in-class performance and the accounts found in their stories, the present author considers this approach effective in assisting students throughout the course. Specifically, the course allows students to acquire requisite initial knowledge to be able to analyse, interpret, and present quantitative data that is needed for their research projects. Additionally, course procedures help the instructor to monitor and support his/her students in real-time and to encourage continual reflection and exploration practices which are important to develop expertise. Importantly, however, to provide a more rigorous evaluation of the effectiveness of the training, quantitative assessments of student learning outcomes are needed, indicating a potential area for future research.

Overall, this methods tutorial contributes to existing methodological literature by providing guidance on teaching quantitative data analysis with the open-source software, GNU PSPP. The approach described in the paper appears to have the potential to foster research independence among students, faculty, and researchers, which is crucial for their future research projects. Notably, the adaptability of this approach allows for its application to the training of analytic skills in the fields beyond applied linguistics, including STEM and social sciences. Lastly, the use of GNU PSPP, an open-source alternative to commercial statistical software, may be relevant for higher education institutions, particularly those with financial constraints that cannot afford the price of software subscriptions, as they develop and implement training programmes.

References

Bedyńska, S. and M. Cypryańska. 2013. (eds.). *Statystyczny drogowskaz 1. Praktyczne wprowadzenie do wnioskowania statystycznego*. Warszawa: Wydawnictwo Akademickie Sedno.

Collins, A., et al. 1987. "Cognitive Apprenticeship: Teaching the Craft of Reading, Writing, and Mathematics. Technical Report No. 403." Retrieved from https://files.eric.ed.gov/fulltext/ED284181.pdf. 20 Aug. 2024.

Collins, A., et al. 1991. "Cognitive Apprenticeship: Making Thinking Visible." *American Educator* 15.3. 6–11.

Collins, A. and M. Kapur. 2014."Cognitive Apprenticeship." In: Sawyer, R. K. (ed.). *The Cambridge Handbook of the Learning Sciences* 291. Cambridge: Cambridge UP, 109–27.

Creswell, J. W. 2012. *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research*. Harlow: Pearson.

Dörnyei, Z. 2007. *Research Methods in Applied Linguistics*. Oxford: Oxford UP.

Gass, S., et al. 2021. "Coming of Age: The Past, Present, and Future of Quantitative SLA Research." *Language Teaching* 54.2. 245–58.

Gönülal, T. 2020. "Statistical Knowledge and Training in Second Language Acquisition: The Case of Doctoral Students." *ITL-International Journal of Applied Linguistics* 171.1. 62–89.

Gönülal, T., S. Loewen and L. Plonsky. 2017. "The Development of Statistical Literacy in Applied Linguistics Graduate Students." *ITL-International Journal of Applied Linguistics* 168.1. 4–32.

Larson-Hall, J. 2015. *A Guide to Doing Statistics in Second Language Research Using SPSS and R*. London: Routledge.

—. 2017. "Moving beyond the Bar Plot and the Line Graph to Create Informative and Attractive Graphics 1." *The Modern Language Journal* 101.1. 244–70.

Larson-Hall, J. and R. Herrington. 2010. "Improving Data Analysis in Second Language Acquisition by Utilizing Modern Developments in Applied Statistics." *Applied Linguistics* 31.3. 368–90.

Lazaraton, A., et al. 1987. "Forming a Discipline: Applied Linguists" Literacy in Research Methodology and Statistics." *TESOL Quarterly* 21.2. 263–77.

Li, S. and C. Seale. 2008. "Learning to Do Qualitative Data Analysis: An Observational Study of Doctoral Work." *Qualitative Health Research* 17. 1442–52. https://doi.org/10.1177/1049732307306924.

Lindstromberg, S. 2016. "Inferential Statistics in Language Teaching Research: A Review and Ways Forward." *Language Teaching Research* 20.6. 741–68. https://doi.org/10.1177/1362168816649979.

Liu, H., et al. 2017. "Pedagogical Beliefs and Attitudes toward Information and Communication Technology: A Survey of Teachers of English as a Foreign Language in China." *Computer Assisted Language Learning* 30.8. 745–65. https://doi.org/10.1080/09588221.2017.1347572.

Loewen, S., et al. 2014. "Statistical Literacy among Applied Linguists and Second Language Acquisition Researchers." *TESOL Quarterly* 48.2. 360–88.

Loewen, S., et al. 2020. "How Knowledgeable Are Applied Linguistics and SLA Researchers About Basic Statistics? Data from North America and Europe." *Studies in Second Language Acquisition*. 42.4. 871–90. https://doi.org/10.1017/S0272263119000548.

Lowie, W. and B. Seton. 2012. *Essential Statistics for Applied Linguistics*. New York: Palgrave Macmillan.

Mackey, A. and S. M. Gass. 2005. *Second Language Research: Methodology and Design*. Mahwah: Lawrence Erlbaum.

Matsuo, M. and T. Tsukube. 2020. "A Review on Cognitive Apprenticeship in Educational Research: Application for Management Education." *The International Journal of Management Education* 18.3. 100417. https://doi.org/10.1016/j.ijme.2020.100417.

McKinley, J. and H. Rose. 2020. (eds.). *The Routledge Handbook of Research Methods in Applied Linguistics*. Abingdon: Routledge.

Norris, J. M. 2015. "Statistical Significance Testing in Second Language Research: Basic Problems and Suggestions for Reform." *Language Learning* 65.1. 97–126.

Norris, J. M., et al. 2015. "Guidelines for Reporting Quantitative Methods and Results in Primary Research." *Language Learning* 65.2. 470–76.

Phakiti, A., et al. 2018. (eds.). *The Palgrave Handbook of Applied Linguistics Research Methodology*. London: Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59900-1.

Pitura, J. 2022. "Using the E-Questionnaire in Qualitative Applied Linguistics Research." *Research Methods in Applied Linguistics* 2.1. 100034. https://doi.org/10.1016/j.rmal.2022.100034.

—. 2023. "Revealing Attitude toward Statistics among MA TEFL Students: A Systemic Functional Linguistics Perspective." *Beyond Philology: An International Journal of Linguistics, Literary Studies and English Language Teaching* 20.1. 115–56.

Plonsky, L. 2015. *Advancing Quantitative Methods in Second Language Research*. New York: Routledge.

Plonsky, L. and F. L. Oswald. 2017. "Multiple Regression as a Flexible Alternative to ANOVA in L2 Research." *Studies in Second Language Acquisition* 39.3. 579–92.

Poland, Ministry of Science and Higher Education. *W sprawie standardu kształcenia przygotowującego do wykonywania zawodu nauczyciela*. Dz.U. 2019 poz. 1450, 25 July 2019.

Saadati, F., et al. 2015. "Effect of Internet-Based Cognitive Apprenticeship Model (*i*-CAM) on Statistics Learning among Postgraduate Students." *PLOS ONE* 10.7. e0129938. https://doi.org/10.1371/journal.pone.0129938.

Wild, C. J. and M. Pfannkuch. 1999. "Statistical Thinking in Empirical Enquiry." *International Statistical Review* 67.3. 223–48. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x.