



Exploring task- and student-related factors in the method of propositional manipulation (MPM)

[Jimmie Leppink](#)

[Nick J. Broers](#)

[Tjaart Imbos](#)

[Cees P. M. van der Vleuten](#)

[Martijn P. F. Berger](#)

Maastricht University

Journal of Statistics Education Volume 19, Number 1 (2011),

www.amstat.org/publications/jse/v19n1/leppink.pdf

Copyright © 2011 by Jimmie Leppink, Nick J. Broers, Tjaart Imbos, Cees P. M. van der Vleuten and Martijn P. F. Berger all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Statistics education; Method of propositional manipulation (MPM); Conceptual understanding; Guided self-explanation; Cognitive load.

Abstract

The method of propositional manipulation (MPM) aims to help students develop conceptual understanding of statistics by guiding them into self-explaining propositions. To explore task- and student-related factors influencing students' ability to learn from MPM, twenty undergraduate students performed six learning tasks while thinking aloud. The results indicate that whether students learn from MPM depends on their statistics proficiency level, the subject matter, the number of propositions in the learning task, and the instructions. MPM learning tasks should be tailored to the students' level of expertise and students should be instructed more than once to integrate all propositions in the learning task into their arguments.

1. Introduction

The statistics knowledge domain comprises abstract concepts that frequently build on other concepts and have no meaning outside the domain. This, together with other factors (e.g., the place in the study curriculum, the student's background and motivation, inappropriate instructional formats), makes it difficult for students to develop conceptual understanding of statistics (i.e., an understanding of the statistical concepts and the relationships between these

concepts; [Huberty, Dresden, & Bak, 1993](#)). The method of propositional manipulation, in short MPM ([Broers, 2008](#)), aims to help students develop conceptual understanding by guiding them into self-explaining the subject matter.

1.1 Theoretical groundwork for MPM

When studying statistics literature, attending a lecture, or when performing a learning task on statistics, students are confronted with important concepts and core ideas. Students first have to isolate the important ideas by deriving their constituent elements, and then relate and integrate these elements into schemata and gradually develop an integrated knowledge network ([Novak, 2002](#)). Knowledge of isolated statistical ideas and elements is called *propositional knowledge*, whereas the ability to relate and integrate these elements is called *conceptual understanding* ([Huberty et al., 1993](#)). Propositional knowledge is a necessary but not sufficient condition for conceptual understanding ([Marshall, 1995](#)). Developing conceptual understanding also involves *self-explanation* and *argumentation* ([Aleven & Koedinger, 2002](#); [Fischer, 2002](#); [Knipfer, Mayr, Zahn, Schwan, & Hesse, 2009](#)). In the domain of statistics, guiding students into self-explanation as in MPM appears to enhance learning outcomes more than unguided self-explanation ([Broers & Imbos, 2005](#); [Broers, Mur, & Bude, 2005](#)), most likely because unguided self-explanation requires students to find out themselves which are the relevant propositions in the subject matter. The latter can easily lead to disorientation on the part of the students.

Learning imposes cognitive load on students ([Van Merriënboer & Sweller, 2005](#)). Cognitive load consists of three types of load that are assumed to be additive: intrinsic load, germane load, and extraneous load. Intrinsic load depends on task complexity and the students' statistics proficiency level. This type of load should be manipulated in instructional design by selecting learning tasks that fit to the students' statistics proficiency level ([Schnitz & Kuerschner, 2007](#)). As the intrinsic load imposed on students when studying statistics is usually high, a learning task that is too difficult will easily lead to cognitive overload ([Kalyuga, 2009](#)). Furthermore, all instructional features not directly beneficial for learning impose extraneous load on the student. To have sufficient capacity available for germane load, that is load from instructional features and learning processes enhancing learning (e.g., self-explanation and argumentation), extraneous load should be minimized. Germane load is not only constrained by intrinsic and extraneous load, but also by students' interests and learning orientations, and affective and motivational aspects.

Having students learn by themselves in the domain of statistics easily leads to cognitive overload and disorientation on the part of the students, and as a consequence they will not develop proper knowledge and understanding of the subject matter. There is a need for an instructional format that stimulates the student to self-explain without experiencing cognitive overload, and this is exactly the focus of MPM.

1.2 MPM and domain-specific thought-processes in statistics

MPM comprises three steps. In the first step, the instructor determines the subject matter and divides it into a limited number of propositions. Propositions are statements referring to single statistical ideas and concepts (e.g., arithmetic mean, mode, and z -score). The number of

propositions depends on size and content of the subject matter. As mentioned previously, intrinsic cognitive load needs to be manipulated in the instructional design by selecting learning tasks matching the students' statistics proficiency level. Therefore, which propositions are chosen by the instructor should depend on the students' statistics proficiency level. The instructor formulates questions, each referring to one proposition. Examples of propositions and questions referring to these propositions are presented in [Box 1](#).

Box 1. Example of propositions and questions referring to propositions

Proposition 1: a z-score expresses the deviation of a score from the arithmetic mean, relative to the standard deviation.

Question referring to proposition 1: what is expressed by a z-score?

Proposition 2: the z-score of a score equal to the arithmetic mean equals zero.

Question referring to proposition 2: what is the case when z equals zero?

Proposition 3: the arithmetic mean is strongly influenced by scores in the tail of a skewed distribution.

Question referring to proposition 3: why is the arithmetic mean not robust against skewness?

Proposition 4: the mode in a unimodal distribution is the peak of that distribution.

Question referring to proposition 4: what is expressed by the mode in a unimodal distribution?

Thus, the idea is that if the instructor wants students to learn the four propositions presented in [Box 1](#), (s)he has to formulate questions referring to each of these propositions. By having the instructor determine and decompose the subject matter this way, students have more cognitive resources available for learning. Having the student search for the relevant propositions themselves would increase extraneous load, as this search process is not directly beneficial for learning.

In the second step of MPM, students are instructed to answer the questions formulated in the first step. Students are provided with the questions (e.g., "what is expressed by a z-score?"), not the actual propositions (e.g., "a z-score expresses the deviation of a score from the arithmetic mean, relative to the standard deviation"). The propositions are taught to the students in lectures and they can be found in the literature to be studied. By having students answer questions referring to propositions, they become aware of important misconceptions and they develop the propositional knowledge needed for building conceptual understanding. Students are stimulated to self-explain the subject matter and they are guided into this process of self-explanation by means of the questions. Given the abstract and cumulative nature of statistics and the frequent and tough misconceptions students have about the subject matter, this second step of MPM is a necessary step for developing conceptual understanding of statistics.

It is only in the third step of MPM that students begin to develop conceptual understanding, namely by performing a series of MPM learning tasks. In an MPM learning task, students have to relate and integrate a number of propositions into an argument that proves a given hypothesis to be either true or false. In contrast to propositions, the hypothesis typically comprises multiple statistical ideas and concepts. Therefore, hypotheses are generally of a higher complexity level than propositions. Once students master the propositions (i.e., propositional knowledge), they should relate and integrate these propositions into an argument in such a way that the argument enables them to understand why the hypothesis is true or false (i.e., conceptual understanding).

The propositions have been chosen by the instructor in the first step. For each proposition, the instructor formulates one question (for examples see [Box 1](#)). In the second step, students answer these questions (and thereby discover the propositions). In the third step, the instructor gives the students a hypothesis; then the instructor attaches several propositions formed as questions to this hypothesis. The complexity level and the exact formulation of the hypothesis depend on the learning goals of the statistics course: what interrelationships between statistical ideas and concepts do we want students to know at the end of the course? Next, which questions are attached to the hypothesis depends on the learning goals of the statistics course as well as on the specific content of the hypothesis. Consider the example presented in [Box 2](#).

Box 2. Example of an MPM learning task in the statistics domain

Hypothesis: *If a distribution is unimodal and skewed to the right, the mode of that distribution has a negative z-score.*

- [1] Why is the arithmetic mean not robust against skewness?
- [2] What is expressed by the mode in a unimodal distribution?
- [3] What is expressed by a z-score?
- [4] What is the case when z equals zero?

Suppose, the instructor wants students to learn that although in a unimodal and symmetric distribution the mode and the arithmetic mean are equal, in a unimodal but skewed distribution the arithmetic mean is shifted towards the tail of that distribution. In a unimodal and symmetric distribution, the z -score of the mode equals zero, whereas in a unimodal but skewed distribution the z -score of the mode is not equal to zero. Thus, the instructor formulates the hypothesis presented in [Box 2](#). The hypothesis includes the concepts ‘mode’, ‘unimodal distribution’, and ‘z-score’. This explains why questions [2] and [3] have been attached. Further, the z -distribution is a unimodal and symmetric distribution, meaning that mode, median, and arithmetic mean are equal. This is why questions [1] and [4] are useful here. When students are confronted with the hypothesis only, they may not be able to answer that the hypothesis in question is true. And even if answering this question, the answer – “true” or “false” – may reflect a rule that was learnt by heart right before the exam, without learning the meaning of the statistical concepts the hypothesis comprises. The latter being the case, it is very likely that students will not be able to solve other hypotheses comprising the same propositions.

MPM stimulates students to engage in meaningful learning, as it stimulates them to self-explain the elements underlying the more complex hypothesis. The students must form an argument for

the truth or falsity of the hypothesis based on the answers to the questions and the connections between them. Students are not expected to learn the propositions through MPM learning tasks. It is in the previous (i.e., second) step of MPM that the propositions are presented to the students, in lectures as well as in the (course) literature. Besides, questions should be formulated in such a way that they require only short answers and that each question can be related to at least one other question. Given q number of questions, the argument can comprise a maximum of $q(q - 1)/2$ pairwise connections. The number of valid connections depends on the exact content of the propositions formed as questions. In the example presented in [Box 2](#), students have to create an argument comprising the answers to the four questions, meaning at least three connections and at most six connections. In an MPM learning task, students do not receive instruction on which connections should be made and which connections should be left out. The only instruction students receive is to create their argument in such a way that each question is related to at least one other question, and there is no further instruction around the learning task. A correct argument for the example is displayed in [Box 3](#).

Box 3. Example of an MPM argument in the statistics domain

In a unimodal distribution, the mode is the peak of the distribution [question 2]. In the case of a skewed distribution, the arithmetic mean is strongly influenced by scores in the tail of that distribution [question 1]. Therefore, the arithmetic mean in this distribution will be lying more towards the tail than the mode [questions 1 and 2 are related]. A z-score expresses how many standard deviations the original observation deviates from the arithmetic mean and in which direction [question 3]. In the case that z equals zero, the original observation does not deviate from the arithmetic mean [questions 3 and 4 related]. Given that the distribution here is skewed to the right, the mode is lower than the arithmetic mean, and therefore, the z-score of the mode is negative [questions 2 and 3]. Thus, the hypothesis is correct.

Each question refers to a single statistical idea or concept. Therefore, a correct argument comprises a relevant set of true propositions and can prove a given hypothesis to be either true or false. Which hypothesis and which questions one chooses for constructing an MPM learning task depends on the learning goals of the statistics course as well as on the students' statistics proficiency level. For example, the learning task presented in [Box 2](#) may increase understanding on the part of students who have just attended the relevant lecture and studied the accompanying literature, whereas for someone who has profound knowledge and understanding of descriptive statistics, this learning task may be too easy to increase understanding. Thus, when formulating the hypothesis and questions in an MPM learning task, students' statistics proficiency level has to be taken into account. Further, to put a learning task into a (real-life) context, contextual information (e.g., a problem) can precede the hypothesis.

Developing conceptual understanding requires a sound propositional knowledge as well as self-explanation and argumentation, and MPM encompasses all these elements. By having the instructor choose the propositions, students are guided into self-explanation of these propositions, which helps them to develop propositional knowledge. Next, it is the manipulation

of propositions in learning tasks that guides students into self-explanation and argumentation on a higher, more complex level, which helps them to develop conceptual understanding.

1.3 Factors affecting MPM success

There are at least five factors that can affect students' ability to perform an MPM learning task and learn from such a task.

First of all, lack of propositional knowledge may hamper students' ability to create an argument. As mentioned before, studying and self-explaining the propositions is a necessary step towards developing conceptual understanding. A question of interest to the current study is to what extent instructing students to study and self-explain the propositions (i.e., in the form of questions) helps them to develop propositional knowledge.

Second, even if students have the propositional knowledge needed to create their argument, there is no guarantee that creating the argument contributes to learning, and whether this interacts with students' statistics proficiency level.

Third, an interesting question is whether students use all their propositional knowledge in the argument explicitly, or whether they tend to leave some propositions implicit.

Fourth, as in the argument every question needs to be related to at least one other proposition, choosing more propositions means students have to produce more relationships. It can be expected that cognitive load increases as the number of propositions to be integrated into the argument increases. An interesting question is the consequences of choosing more propositions in terms of learning outcomes.

Fifth, depending on students' statistics proficiency level and on the complexity of the subject matter, increased cognitive load can either increase or decrease learning outcomes. Proficient students may learn optimally from an MPM learning task on relatively complex subject matter, for example from inferential statistics, or from a learning task comprising a higher number of propositions, whereas less proficient students may only benefit from an MPM learning task on less complex subject matter, for example from descriptive statistics, or from a learning task consisting of only a few propositions.

The current study addressed the abovementioned factors that might affect MPM success with five research questions:

1. To what extent do students understand the propositions in an MPM learning task after studying the accompanying questions that refer to these propositions?
2. What is the effect of creating an MPM argument on cognitive load and learning outcomes?
3. To what extent do students integrate all propositions, represented by the questions, into their argument?
4. How does the number of propositions in an MPM learning task affect cognitive load and learning outcomes?

5. Is MPM equally effective for relatively complex subject matter (e.g., inferential statistics) as for less complex subject matter (e.g., descriptive statistics)?

Before examining MPM learning tasks in an experimental setup, it is important to have an overall idea about what factors influence a student's ability to learn from such tasks and how these factors interact. To acquire as much knowledge as possible about how an instructional method works in practice, it is important to combine different research methods. Explorative studies may prioritize subsequent experimental studies. Therefore, the current study was explorative, combining quantitative measures for cognitive load and qualitative measures (i.e., a mixed method approach), using a technique from the cognitive research tradition, namely thinking aloud while performing a series of MPM learning tasks.

2. Method

The research questions were studied by having students with different statistics proficiency levels think aloud while working on a total of six learning tasks.

2.1 Participants

Twenty bachelor psychology students who passed the first-year statistics exam volunteered. The first-year statistics course covered probability calculus, sampling distributions, null hypothesis testing, confidence intervals, *t* test, one-way analysis of variance (ANOVA), and χ^2 test. Students with an exam score of six or higher on a ten point scale, pass the exam. Typically, about 55% of the students pass this exam the first time. To create sufficient variability in statistics proficiency level, two samples were drawn: ten students with a low proficiency level (i.e., low proficiency group), and ten students with a high proficiency level (i.e., high proficiency group). The lower proficiency group consisted of students who passed the exam with six or seven on a ten point scale, and the higher proficiency group consisted of students with grades eight or higher.

2.2 Materials

The learning material is found in the [Appendix](#). To answer the questions on the number of propositions and the potential of MPM for descriptive versus inferential statistics, the six learning tasks differed in topic and in the number of propositions to be linked in the argument. Three of the learning tasks covered different topics within descriptive statistics, including association in a two-way table (three propositions), histogram and intervals of scores (four propositions), and the usability of the correlation coefficient judging from a scatterplot (five propositions). The other three learning tasks covered different topics within inferential statistics, namely the expected mean (three propositions), Type II error (four propositions), and statistical significance (five propositions).

To answer the question to what extent students understand the propositions in an MPM learning task after studying the accompanying questions, they received a list with the questions that appeared in these learning tasks, with the instruction to study them carefully from [Moore and McCabe's \(2009\)](#) textbook. Since the questions appeared in the learning tasks to be performed a

week later as well, a separate list of the questions to be studied at home is not included in the [Appendix](#).

As we were interested in effects of different aspects of MPM on learning outcomes as well as on cognitive load, we needed a device to measure cognitive load imposed on the students when performing the learning tasks. Cognitive effort is an accepted indicator for cognitive load imposed on a student ([Paas, 1992](#)). Typically, students have to indicate on a nine point scale how much cognitive effort performing a task or solving a problem required from them, one being virtually no cognitive effort at all and nine being the maximum cognitive effort. Although most studies use the nine point scale, to create more variability in cognitive effort we used a visual analog scale (VAS). In the latter, students are instructed to indicate how much cognitive effort performing a task or solving a problem required by drawing a small vertical line on a horizontal continuous line going from 0 (left) to 100 (right). Such scales have been used in numerous studies in various domains the last nineteen years, and have acceptable reliability and validity for those domains (e.g., [DeLoach, Higgins, Caplan, & Stiff, 1998](#); [Gallagher, Bijur, Latimer, & Silver, 2002](#); [Myles, Troedel, Boquest, & Reeves, 1999](#)). To our knowledge, reliability and validity estimates for studies in educational settings are still lacking. The reason why we chose the VAS in the current study was to create more variability in cognitive effort ratings. For each learning task, the VAS was administered twice. In the next section (procedure), it is explained why and how this was done.

2.3 Procedure

For all the learning tasks, the procedure was the following. To make sure that students understood what they had to do in a learning task, they had to first read and summarize aloud contextual information and a hypothesis. Eventual misreading or misinterpretation of language could be corrected by the researcher. At this point, students did not yet receive the questions (referring to the propositions) to be linked into an argument. Students were instructed to explain whether the hypothesis was true or false, using the contextual information at hand. Once they had given their explanation, they indicated on the VAS how much cognitive effort they experienced while performing this task. At this point, we had one solution and one cognitive effort indication (for every student) for the learning task in question. In each learning task, this solution and cognitive effort indication served as baseline measurement.

After this baseline measurement, students were confronted with the underlying questions in the learning task, each referring to one proposition. Before instructing students to create an MPM argument, students were asked to answer each of the questions presented to them (i.e., three, four, or five questions, depending on the learning task). Their answers provided information on the question to what extent students understand the propositions in an MPM learning task after studying the accompanying questions that refer to these propositions. As expected, many students still demonstrated incomplete knowledge of some of the propositions. Since we wanted to know whether students having sufficient propositional knowledge can create an MPM argument – propositional knowledge is a necessary condition for conceptual understanding – we provided all students (i.e., including those who demonstrated complete knowledge) with standard answers to the questions referring to the propositions and instructed them to create an argument integrating these propositions (i.e., answers). Their argument should prove whether the

hypothesis in the learning task was true or false. Once they had created their argument, they indicated again on the VAS how much cognitive effort was experienced while performing this task. In each learning task, this solution and cognitive effort indication served as the after-treatment measurement.

Thus, in each learning task we had two solutions for every student as well as two cognitive effort indications, one after their explanation without explicitly referring to the propositions and the other after they had created their MPM argument. This enabled us to estimate the effect of creating an MPM argument on learning outcomes and cognitive load. If MPM was successful in these learning tasks, students' MPM arguments would comprise more information than their explanations given some minutes earlier. With regard to cognitive load, we did not have specific expectations, partly because of the diversity in learning tasks.

Apart from the comparison of students' MPM arguments with their explanations given before the confrontation with the questions, the MPM arguments provided information with regard to the question to what extent students explicate their knowledge of the propositions (i.e., answers to the questions) and relationships between them in their argument. To acquire additional information on the latter, every pair of questions was isolated and students had to explain what relationship they thought existed between the two questions.

2.4 Data analysis

The current study combined quantitative measures (i.e., with regard to cognitive load) and qualitative measures (i.e., think-aloud protocols transcribed verbatim). For the qualitative analyses, two researchers rated independently from each other. Differences in interpretation were discussed in order to seek consensus.

2.4.1 The effect of studying propositions

For the first research question, to what extent students understand the propositions in an MPM learning task after studying the accompanying questions that refer to these propositions, students' answers to the questions in the learning tasks were compared with the answers derived from [Moore and McCabe's \(2009\)](#) textbook and coded either correct or incorrect by two independent coders. Initial agreement between the coders was high (Cohen's $\kappa = .90$).

2.4.2 The effect of creating an MPM argument

For the second research question, on the effect of creating an MPM argument on cognitive load and learning outcomes, students' MPM arguments were compared with their explanations given before the confrontation with the questions. Two independent raters rated each argument as: (a) a correct and complete argument leading to a correct hypothesis evaluation (i.e., a correct "true" or "false"), (b) an incomplete (i.e., not all propositions integrated explicitly) and/or partly incorrect argument leading to a correct hypothesis evaluation, or (c) an incomplete and/or (partly) incorrect argument leading to an incorrect hypothesis evaluation (i.e., an incorrect "true" or "false"). Initial agreement between the coders was also high (Cohen's $\kappa = .81$).

Further, as in every learning task students indicated twice how much cognitive effort the learning task required from them (i.e., once before and once after creating their MPM argument), split-plot analysis of variance (ANOVA) was performed for comparing the proficiency groups with regard to the effect of creating an MPM argument on cognitive load. The level of significance α was Bonferroni corrected for each of the three p -values (i.e., one with regard to the interaction effect, and two with regard to the main effects) to correct for multiple testing and increased overall Type I error probability.

2.4.3 The extent to which students explicitly integrate propositions in their argument

The coding of the MPM arguments also provided information on the question to what extent students explicitly integrate propositions in their argument. However, the question that arises here is whether or not mentioning a particular relationship between two propositions reflects a mere tendency to leave out a relationship that is known by the student, or it reflects a lack of knowledge of that particular relationship on the part of the student. Therefore, it was determined per pair of propositions whether students indicated a correct relationship between the two propositions. Given that students had the answers to the questions referring to the propositions it was relatively easy for them to provide comments on the relationships.

2.4.4 The number of propositions in an MPM learning task

The effect of the number of propositions in a learning task on learning outcomes was examined by using the analyses of students' arguments described in section 2.4.2. For the effect on cognitive load, two-way within-subjects ANOVA was performed. Factors were the number of propositions and measurement point (i.e., baseline measurement being after students' explanation without the propositions to be integrated, after-treatment measurement being after students' MPM argument). The level of significance α was Bonferroni corrected for each of the three p -values.

2.4.5 Descriptive statistics and inferential statistics

Similar to the question on the effect of the number of propositions in a learning task, the question on the potential of MPM for descriptive versus inferential statistics was examined by using the analyses of students' arguments described in section 2.4.2. For the effect on cognitive load, two-way within-subjects ANOVAs were performed. Factors were subject matter (i.e., descriptive versus inferential statistics) and measurement point (i.e., baseline measurement being after students' explanation without the propositions to be integrated, after-treatment measurement being after students' MPM argument). The level of significance α was Bonferroni corrected for each of the three p -values.

3. Results

Each of the following paragraphs addresses the results with regard to one of the research questions of the current study.

3.1 The effect of studying propositions

The data suggest that students find it difficult to describe abstract statistical concepts. Although all students could answer most questions correctly and demonstrated partial knowledge with regard to those questions they could not answer correctly, a total of three questions had a very low number of correct answers in both groups. First of all, a total of fifteen students did not mention that the correlation coefficient is about *linear* relationship and not about just any relationship. Second, students' descriptions of the *p*-value revealed that many students find it difficult to interpret this value as a conditional probability. Third, although students knew examples of test statistics, they could not give a general definition of this term. Finally, eight of the ten students in the lower proficiency group confused the sampling distribution with the distribution of sample scores. In the higher proficiency group, five students made this mistake.

3.2 The effect of creating an MPM argument

With regard to the effect of creating an MPM argument, the data can be summarized as follows. First, when asking students to evaluate hypotheses they consider easy – because the hypothesis is self-evident, a rule learned by heart, or too easy for their statistics proficiency level – they hardly motivate their evaluations. Second, creating an MPM argument does not guarantee that students replace their misconceptions by correct knowledge. A quote from a student in the lower proficiency group:

“...eh, yes the value is close to 0 and thus I would say that there is almost no association, *linear* association, and, yes, I cannot see any non-linear association here, because whether you draw a straight line or any other line you will not manage because the points are so spread out and thus I suppose that the correlation coefficient, eh, yes, does give the correct value and given that this is here .03, it is, eh, indeed not a good summary of the association between *x* and *y*”

Third, given the difficulties students experience when describing the statistical concepts mentioned, creating an MPM argument integrating these concepts is likely to be difficult for them as well. An example is the following:

“Okay, so the hypothesis was eh, whether the expected mean can be expressed in a number... hmm... now I am confused... eh... okay... so, I have the, eh, mean of the, eh, neuroticism scores... hmm, so the distribution of the sample is *not* the same as the sampling distribution... hmm, yes, but the distribution of the sample... sample mean is eh... I just said that the hypothesis is true and I still think that eh... well the sample mean is the expected mean, or not? The, sorry, eh, if the sample is drawn at random, then they must be equal... [researcher reminds the student that *all propositions have to be integrated into the argument, conform the instructions*] ... hmm, yes, I, I see no relation here between the questions... I have only the distribution of the sample and not the sampling distribution... so eh... that cannot be?... eh but then the hypothesis is incorrect... [researcher reminds the student that *all propositions have to be integrated into the argument, conform the instructions*] ... hmm, yes, here I have only the mean of the sample, so, eh, that is just a one value of... eh, it is just one sample... [researcher reminds the student that *all propositions have to be integrated into the argument, conform the instructions*] eh, the sample mean is not equal to the expected

mean, and I do not have the population mean here... yes... the population mean equals the expected mean... so the hypothesis is false, because the population mean is not given and that one is equal to the expected mean, now I cannot compute the population mean, because I just have the distribution of one sample, and thus not the sampling distribution of the mean when you repeat an infinite number of times, because, eh, de expected mean is the mean of the sampling distribution and that one is not given here... [researcher: *thus the hypothesis is?*] ... false."

This example is from a student in the higher proficiency group. Students' arguments in this learning task illustrate that frequent misconceptions about the expected mean are that the expected mean equals the sample mean or that the expected mean can be computed from mere sample data. Before the instruction to create an MPM argument in this learning task, sixteen of the twenty students – eight in each group – gave an incorrect explanation leading to an incorrect hypothesis evaluation. The instruction to create an MPM argument made six of the proficient students aware of their mistake, and as a result they came to a correct hypothesis evaluation. In the lower proficiency group, this change was limited to one student. This finding is in line with the finding reported in the previous section that describing the concept of sampling distribution is difficult, especially for the less proficient students.

With regard to cognitive load, on average, proficient students reported lower cognitive load than their less proficient peers. [Table 1](#) displays the average cognitive efforts and standard deviations for both proficiency groups before and after the instruction to create an MPM argument, for the three learning tasks on descriptive statistics.

Table 1

Means (and SD) of cognitive effort required for the learning tasks on descriptive statistics.

Condition	N	M (SD)
Explanation before the questions		
Lower proficiency group	10	45.01 (17.73)
Higher proficiency group	10	31.79 (17.39)
MPM argument		
Lower proficiency group	10	39.40 (13.41)
Higher proficiency group	10	31.52 (16.58)

The group by condition interaction was not significant, $[F(1, 18) = 0.336, p > .50, \eta^2 = .018]$ and the same was the case for the main effect of condition, $[F(1, 18) = 0.408, p > .50, \eta^2 = .022]$, as well as for the group effect, $[F(1, 18) = 3.43, p = .08, \eta^2 = .160]$. As the effect sizes indicate a large size effect for the group effect, absence of statistical significance is probably due to small sample sizes. [Table 2](#) displays the average cognitive efforts and standard deviations for both proficiency groups before and after the instruction to create an MPM argument, for the three learning tasks on inferential statistics.

Table 2

Means (and SD) of cognitive effort required for the learning tasks on inferential statistics.

Condition	N	M (SD)
Explanation before the questions		
Lower proficiency group	10	45.96 (12.00)
Higher proficiency group	10	36.34 (17.35)
MPM argument		
Lower proficiency group	10	54.41 (13.41)
Higher proficiency group	10	46.23 (16.40)

The group by condition interaction was not significant, $[F(1, 18) = 0.048, p > .80, \eta^2 = .003]$, the main effect of condition was significant, $[F(1, 18) = 7.729, p < .05, \eta^2 = .300]$, and the group effect was not significant, $[F(1, 18) = 2.340, p > .10, \eta^2 = .115]$. As the effect sizes indicate a medium to large size effect for the group effect, absence of statistical significance is probably due to small sample sizes.

The data suggest that evaluating a hypothesis by means of an MPM argument imposes additional cognitive load on students when learning inferential statistics, but not when learning descriptive statistics.

3.3 The extent to which students explicitly integrate propositions in their argument

Despite the instruction to explicitly integrate propositions, students tend to restrict themselves to merely listing the propositions. In most cases, it was only after repeated asking by the researcher to relate and integrate propositions that students attempted to do so. The researchers anticipated this possibility, and therefore after creating the MPM argument, students were instructed per pair of underlying propositions what the relationship between the two propositions is. Although initially not all students were able to give appropriate answers to all questions, they were now able to describe the relationships for nearly all pairs of propositions. Not clear is whether this ability was a consequence of seeing the standard answers to the questions or partly an effect of creating an MPM argument. These findings and those in the previous two sections together suggest two practical implications for MPM. First, sufficient propositional knowledge is a firm necessary condition for creating an MPM argument. Second, to avoid that relevant propositional knowledge does not appear in the argument, the instruction to integrate all propositions in the learning task in the argument should be clear and repeated.

3.4 The number of propositions in an MPM learning task

[Table 3](#) presents the average cognitive efforts and standard deviations for the number of propositions (i.e., for both contents and for all twenty students) before and after the instruction to create an MPM argument.

Table 3

Means (and SD) of cognitive effort experienced as a function of the number of propositions.

Condition	<i>N</i>	<i>M</i> (SD)
Explanation before the questions		
Three propositions	20	49.03 (19.21)
Four propositions	20	37.90 (16.75)
Five propositions	20	32.40 (14.57)
MPM argument		
Three propositions	20	55.41 (13.27)
Four propositions	20	34.76 (21.06)
Five propositions	20	38.51 (17.82)

Two-way within-subjects ANOVA revealed a non-significant interaction between the number of propositions and the learning format, $[F(2, 38) = 2.405, p > .10, \eta^2 = .112]$, and a significant main effect for the number of propositions, $[F(2, 38) = 15.552, p < .001, \eta^2 = .450]$. From further analysis it appears that the proposition effect is quadratic, $[F(1, 18) = 6.229, p < .05, \eta^2 = .247]$ and that a learning task consisting of three underlying propositions requires significantly more cognitive effort than a learning task consisting of four or five underlying propositions. It is difficult to interpret this quadratic effect, as it may indicate some kind of novelty effect: for both descriptive and inferential statistics, the first learning task students were confronted with consisted of three propositions and the learning tasks following consisted of four and five propositions respectively.

3.5. Descriptive statistics and inferential statistics

[Table 4](#) presents the average cognitive efforts and standard deviations for statistical topic before and after the instruction to create an MPM argument.

Table 4

Means (and SD) of cognitive effort required as a function of statistical topic.

Condition	N	M (SD)
Explanation before the questions		
Descriptive statistics	20	38.40 (18.39)
Inferential statistics	20	41.15 (15.23)
MPM argument		
Descriptive statistics	20	35.46 (15.33)
Inferential statistics	20	50.32 (15.17)

The average cognitive effort to perform the assignment was not only higher for inferential assignments than for descriptive assignments, a difference in trend for the different subject matters exists as well. Two-way within-subjects ANOVA revealed a significant interaction between content and learning format, $[F(1, 19) = 5.057, p < .05, \eta^2 = .210]$. Subsequent analysis made clear that the difference between conditions was significant for the learning tasks on basic inferential statistics, $[F(1, 19) = 8.140, p < .01, \eta^2 = .300]$, and that creating an MPM argument required significantly more cognitive effort in learning tasks on elementary inferential statistics than in learning tasks on descriptive statistics, $[F(1, 19) = 9.730, p < .05, \eta^2 = .339]$. Apparently, it demands a considerable cognitive effort from the student to do an inferential statistics assignment in MPM format. Together with the finding that, in the higher proficiency group, confronting students in MPM format with the concepts of sampling distribution and expected mean leads to an increased awareness of a misconception and, as a consequence, better performance, one can infer that the additional cognitive effort due to working in MPM format can lead to better learning outcomes, at least for those having a certain proficiency in (inferential) statistics.

4. Discussion

The first major finding is that instructing students to study a list of propositions is effective for the easier propositions but less effective for propositions referring to abstract statistical concepts. The instruction to study the propositions should be such that students spend more time studying abstract statistical concepts. This could be achieved by a list of propositions in which abstract concepts are represented by more propositions than easier concepts. In the current study, each concept was represented by just one proposition. Future studies may examine the effect of more than one proposition referring to an abstract concept. Sufficient propositional knowledge is a necessary condition for creating an MPM argument. Therefore, if representation by abstractness

in the propositions list leads to more propositional knowledge, the quality of MPM arguments may be increased as well.

With regard to the effect of creating an MPM argument, the current study reveals two extremes. On the one hand, creating an MPM argument integrating abstract statistical concepts is difficult and does not necessarily lead to misconception awareness with regard to these concepts. On the other hand, when asking students to evaluate hypotheses they consider easy – because the hypothesis is self-evident, a rule learned by heart, or too easy for their statistics proficiency level – they hardly motivate their evaluations, nor do they feel motivated to create an MPM argument. In line with findings in previous studies ([Kalyuga, Ayres, Chandler, & Sweller, 2003](#); [Wetzel, 2009](#)), the complexity level of each of the aspects of an MPM learning task – contextual information, hypothesis and underlying propositions – must be in accordance with the students' statistics proficiency level. A very easy learning task is not likely to stimulate students to create an MPM argument, whereas a very difficult learning task may lead to cognitive overload for the students. Future studies should focus on the development and validation of MPM learning tasks for different statistics proficiency levels.

Despite the instruction to explicitly integrate propositions, students tend to restrict themselves to merely listing the propositions. Although in some learning tasks, this behavior is most likely the consequence of a discrepancy between the students' statistics proficiency level and task difficulty, future studies could focus on the effectiveness of different forms of the instruction to integrate all propositions in the argument. For example, to avoid omission of relevant propositional knowledge in the argument, one could repeat the instruction every time a link between two propositions is made by the student. Another type of instruction is to formulate more MPM arguments in a particular learning task.

Since we did not counterbalance the order of learning tasks for the students, the findings with regard to the number of propositions in an MPM learning task do not enable us to draw straightforward conclusions, only that it appears important to make the student familiar with MPM. The findings with regard to cognitive effort appear to reflect that once the student is familiar with MPM, the instructional format itself imposes less cognitive load on the student. To examine the effect of the number of propositions in an MPM learning task, future studies should either treat this effect as a between-subjects factor or counterbalance the order of the learning tasks.

The comparison of descriptive statistics and inferential statistics indicates that the topic itself needs to be complex enough to stimulate students to work according to MPM. However, the finding that the students in the current study are more stimulated to create MPM arguments for inferential statistics than for descriptive statistics may reflect the aforementioned discrepancy between the students' statistics proficiency level and task difficulty. In their study curriculum, a course on descriptive statistics precedes the course on inferential statistics and the students were selected based on their exam score for the latter. Therefore, future studies should examine the potential of MPM for learning descriptive statistics among students starting the course on descriptive statistics.

The current study has a few limitations. First of all, the results are based on small sample sizes and depend upon the actual MPM learning tasks created for this study. Different hypotheses based on more fundamental concepts with more basic questions might lead the students to a better or worse understanding than the actual MPM learning tasks. Second, task-specific characteristics are a confounding factor in some of the comparisons made in the current study. Future studies might use assignments that are hierarchical, that is, different assignments on one specific topic only varying in the number of constituent propositions. The number of propositions should then be a between-subjects factor or the order of the learning tasks should be counterbalanced. A third limitation of the current study is that all students were first instructed to explain in their own words whether the hypothesis in the learning task was true or false and subsequently had to create an MPM argument, based on the standard answers to the questions referring to the propositions. It is possible that the learning effect that was found in some of the learning tasks was partly induced by the fact that students had been working some time already on the learning task and that they had full knowledge of the propositions (i.e., in the form of the standard answers). A fourth limitation is that students in the current study received a single prompt to create an MPM argument. The results indicate that this procedure can help students whose proficiency or prior knowledge matches the difficulty level of the learning task to become aware of a misconception and develop conceptual understanding (i.e., by self-explaining and integrating the constituent propositions). Taking the third and fourth limitation of the current study together, in future studies the instructional format (i.e., explaining in their own words and creating an MPM argument) should be a between-subjects factor consisting of three conditions, namely a control condition in which students explain in their own words, an MPM condition in which the student receives a single prompt to create an MPM argument and a third condition in which the student receives a number of prompts to work this way or receives the instruction to try to formulate more MPM arguments for the same learning task.

The students in the current study did not have a stake in the outcome. A possible consequence is that they may not have tried as hard as students would have if they had a stake in the outcome (e.g., a grade on their results). Therefore, a fifth limitation may be students' learning outcomes in our study were lower than learning outcomes in the educational practice. Finally, as the students in the current study were selected on the basis of their performance on the subject matter, they were not novices. On the one hand, given the findings with regard to the students' proficiency in statistics, some studies should contrast the instructional formats for a larger number of tasks, varying in difficulty level, for different levels of proficiency or prior knowledge. On the other hand, since MPM aims to both guide novices into a complex knowledge domain like statistics and help other students to develop a better (conceptual) understanding of the domain, other studies should use novices (i.e., no prior knowledge about the topic whatsoever) as participants in order to determine whether they can profit from MPM as well. In the latter case, the learning tasks to be constructed need to be relatively easy, since the current study has shown that MPM can only be fruitful if the learning tasks match the students' prior knowledge of the topic.

MPM is a concrete instructional method for the statistics knowledge domain. In this article, we present task- and student-related factors influencing students' ability to learn from an MPM learning task (i.e., statistics proficiency level, subject matter, the number of propositions in the learning task, and the instructions). It is now important to examine each of these factors in subsequent experimental studies.

Appendix

Questions and propositions used

Descriptive statistics, learning task 1

See the cross table below for the association between gender and hair color in a population, in which only three colors of hair exist: blond, brown, and black.

	Female	Male
Blond	45	60
Brown	75	100
Black	30	40
Total	150	200

In this population, for both men and women the univariate distribution for hair color is: 30% blond, 50% brown and 20% black.

Hypothesis:

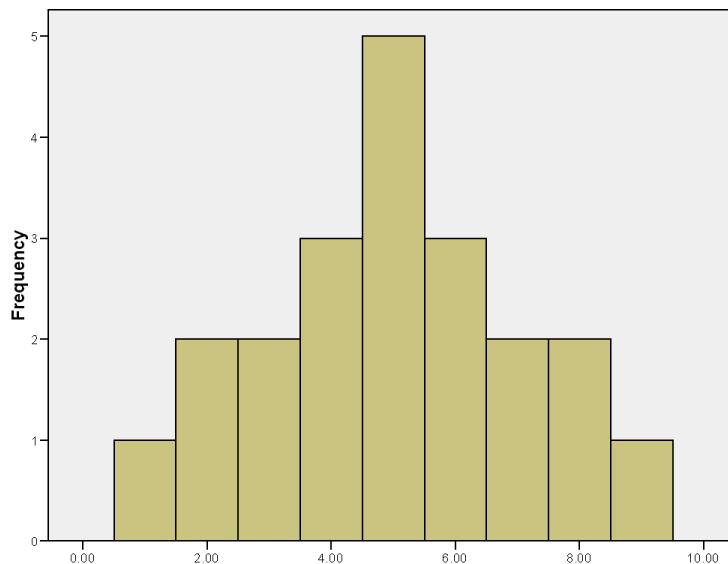
From the information it can be concluded that in this population, there is no association between gender and hair color.

Questions:

- [1] What is the *marginal distribution* of a variable?
- [2] What is the *conditional distribution* of a variable?
- [3] When can we say there is an *association* between two categorical variables?

Descriptive statistics, learning task 2

The histogram below shows the distribution of a variable x . The arithmetic mean of this distribution equals 5 and its standard deviation is 2.12, hence the length of the interval ‘*arithmetic mean plus or minus one time the standard deviation*’ equals 4.24. The first quartile (Q_1) equals 3.5, the third quartile (Q_3) equals 6.5.



Hypothesis:

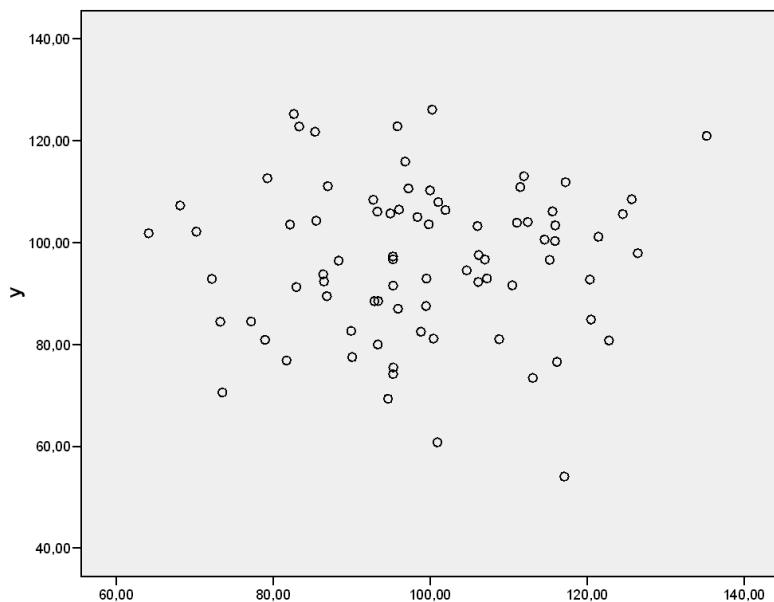
For this distribution of variable x , the interval ‘*arithmetic mean plus or minus one time the standard deviation*’ contains more than 50% of the values.

Questions:

- [1] How is the *interquartile range (IQR)* computed from the values of Q_1 and Q_3 ?
- [2] What range of values is represented by the *IQR*?
- [3] What is the definition of the *median*?
- [4] What can be said about median and arithmetic mean in the case of a *perfectly symmetrical distribution*?

Descriptive statistics, learning task 3

The scatterplot below is about the association between two variables x and y . For this distribution, the correlation coefficient r_{xy} is equal to .03.



Hypothesis:

In this case, the correlation coefficient r_{xy} does not give a good summary of the association between x and y .

Questions:

- [1] What is expressed by the *correlation coefficient* r_{xy} ?
- [2] What values can the correlation coefficient r_{xy} have?
- [3] What is an *outlier* in a scatterplot?
- [4] How can an outlier influence the value of the correlation coefficient r_{xy} ?
- [5] If the association between two variables x and y is *non-linear*, what can be said about the value of the correlation coefficient r_{xy} ?

Basic inferential statistics, learning task 1

Suppose we are interested in the average neuroticism score in a certain population. We draw a random sample of $N = 40$ from that population. Neuroticism was measured on a scale ranging from 0 to 38. The average of the 40 neuroticism scores was equal to 25.

Hypothesis:

The information presented above enables us to express the expected value of the sample mean in a number.

Questions:

- [1] What is a *sampling distribution*?
- [2] What is meant by the *expected mean* of a random variable?
- [3] What parameter equals the expected mean?

Basic inferential statistics, learning task 2

Imagine we test a one-sided hypothesis, our test statistic has a certain value, and the accompanying *P*-value equals .07. Our significance level is .05.

Hypothesis:

The *P*-value of .07 does not give rise to a Type I error, but does give rise to a Type II error.

Questions:

- [1] What is a *P*-value?
- [2] What is meant by *significance level*?
- [3] What is a *Type I error*?
- [4] What is a *Type II error*?

Basic inferential statistics, learning task 3

Hypothesis:

If the value of our test statistic exceeds the critical value, the *P*-value is smaller than the significance level.

Questions:

- [1] What is a *test statistic*?
- [2] What is a *sampling distribution*?
- [3] What is a *P*-value?
- [4] What is meant by *significance level*?
- [5] What is meant by *critical value* in the context of hypothesis testing?

References

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

Broers, N. J., & Imbos, Tj. (2005). Charting and manipulating propositions as methods to promote self-explanation in the study of statistics. *Learning and instruction*, 15, 517-538.

Broers, N. J., Mur, M. C., & Bude, L. (2005). Directed self-explanation in the study of statistics. In G. Burrill & M. Camden (Eds.) *Curricular development in statistics education* (pp. 21-35). Voorburg, The Netherlands: International Statistical Institute.

Broers, N. J. (2008). "Helping students to build a conceptual understanding of elementary statistics," *The American Statistician*, 62, 1-6.

DeLoach, L. J., Higgins, M. S., Caplan, A. B., & Stiff, J. L. (1998). "The visual analogue scale in the immediate postoperative period: intrasubject variability and correlation with a numeric scale." *Anesthesia & Analgesia*, 86, 102-106.

Fischer, F. (2002). Gemeinsame Wissenskonstruktion – Theoretische und methodologische Aspekte [Joint knowledge construction – Theoretical and methodological aspects]. *Psychologische Rundschau*, 53, 119-134.

Gallacher, E., Bijur, P. E., Latimer, C., & Silver, W. (2002). “Reliability and validity of a visual analog scale for acute abdominal pain in the ED.” *American Journal of Emergency Medicine*, 20, 287-290.

Huberty, C. J., Dresden, J., & Bak, B. (1993). “Relations among dimensions of statistical knowledge,” *Educational and Psychological Measurement*, 53, 523-532.

Kalyuga, S. (2009). “Knowledge elaboration: a cognitive load perspective,” *Learning and Instruction*, 19, 402-410.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). “The expertise reversal effect,” *Educational Psychologist*, 38, 23-31.

Knipfer, K., Mayr, E., Zahn, C., Schwan, S., & Hesse, F. W. (2009). Computer support for knowledge communication in science exhibitions: Novel perspectives from research on collaborative learning. *Educational Research Review*, 4, 196-209.

Marshall, S. (1995). *Schemas in problem solving*. Cambridge: Cambridge University Press.

Moore, D. S., & McCabe, G. P. (2009). *Introduction to the practice of statistics* (6th ed.), New York: Freeman.

Myles, P. S., Troedel, S., Boquest, M., & Reeves, M. (1999). “The pain visual analog scale: is it linear or nonlinear?” *Anesthesia & Analgesia*, 89, 1517-1520.

Novak, J. D. (2002). Meaningful learning: the essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education*, 86, 548-571.

Paas, F. (1992). “Training strategies for attaining transfer of problem-solving skills in statistics: a cognitive load approach,” *Journal of Educational Psychology*, 84, 429-434.

Schnitz, W., & Kuerschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19, 469-508.

Van Merriënboer, J., & Sweller, J. (2005). “Cognitive load theory and complex learning: recent developments and future directions,” *Educational Psychology Review*, 17, 147-177.

Wetzels, S. A. J. (2009). “Individualised strategies for prior knowledge activation [dissertation],” Netherlands: Interuniversity Centre for Education Research (ICO).

Jimmie Leppink
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
j.leppink@maastrichtuniversity.nl
Phone: +31 43 388 2279

Nick J. Broers
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
nick.broers@maastrichtuniversity.nl
Phone: +31 43 388 2274

Tjaart Imbos
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
tjaart.imbos@maastrichtuniversity.nl
Phone: +31 43 388 2434

Cees P. M. van der Vleuten
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
c.vandervleuten@maastrichtuniversity.nl
Phone: +31 43 388 5725

Martijn P. F. Berger
Maastricht University
P.O. Box 616, 6200 MD Maastricht, The Netherlands
martijn.berger@maastrichtuniversity.nl
Phone: +31 43 388 2258
