THE TEACHING OF MATHEMATICS
2018, Vol. XXI, 2, pp. 55–72

DATA MODELLING USING INTERACTIVE CHARTS

Djordje M. Kadijevich

Abstract. This paper examines data modelling with interactive charts in upper secondary education. After defining data modelling *per se* and explaining its relationship to mathematical modelling, data modelling with interactive charts is examined in detail, focusing on the main challenges in using them in the context of key data modelling activities. This examination is followed by a report of an empirical pilot study on the difficulty of these activities and the importance of this kind of modelling for upper secondary education. Implications for practice and research are included.

MathEduc Subject Classification: M14

MSC Subject Classification: 97M10

Key words and phrases: Data modelling; interactive charts; upper secondary education.

1. Introduction

We basically use mathematics to describe, predict, or prescribe natural or artificial phenomena (Davis & Hersh, 1990). As a result, mathematical models enable us to understand some aspects of (real-life) problems, to answer specific questions when models are applied, or to undertake actions based on outcomes of the mathematics applied. Based on these uses, we can make a distinction between descriptive and prescriptive modelling: while the former aims at understanding the world modelled, the latter focuses on recommending (well-grounded) actions to change that world. Despite the fact that developing and applying mathematical models to make decisions is quite common in today's business world, the mathematical modelling community has, as Niss (2015) underlined, mostly focused on descriptive modelling.

To improve educational achievements in general, learners' conceptual change (from incomplete to more-or-less complete, professionally/scientifically grounded internal representations) should be fostered though carefully designed modelling activities (Greca & Moreira, 2000). As exemplified in Jonassen (2006), these modelling activities may deal with a number of phenomena (e.g. problems, domain knowledge, systems, and thinking) by using a range of technology-based modelling tools (e.g. hypermedia, spreadsheets, databases, expert systems, and system dynamic tools). Having in mind the growing use of (large) datasets (of financial, scientific, educational, societal, or of other natures), data have become additional phenomena that can be modelled to promote conceptual change (i.e. to learn from data). This can be done by using interactive charts, for example.

According to Lyn English (2016), work with data and uncertainty is central to STEM (science, technology, engineering, and mathematics) education, which would, among other things, support students in making evidence-based claims and decisions regarding societal, economic, environmental, and other issues. It is thus not surprising that there has been noticeable recent demand for the introduction of data modelling into secondary education (e.g. Davison, 2015; Engel, 2017; Gould et al., 2016; Metz, 2015; Ridgway, 2016). This kind of modelling seems to represent an important goal for upper secondary education because it is expected that many students will do some kind of basic data modelling (e.g. examining data by using interactive charts) in their future jobs.

In the mathematical modelling community, data modelling has been mostly studied within primary education in a traditional paper-and-pencil learning environment, as done by English (2012, 2014), for example. As a consequence of this research orientation and the above-mentioned demand, the research presented in this paper examined data modelling in upper secondary education. It was done by using interactive charts, whose application, especially for dashboards (i.e. sets of such charts), has increased considerably in recent years (to view a gallery of dashboards concerning various industries and areas, visit `https://www.idashboards.com/dashboard-examples/`, for example).

In the rest of this paper I will first explain my approach to data modelling and clarify its relationship to mathematical modelling and statistical analysis, having in mind both prescriptive and descriptive modelling. After that, with respect to key data modelling activities, I will examine data modelling with interactive charts, focusing on the main challenges in using them. This examination is followed by a report of an empirical pilot study concerning the difficulty of these activities and the importance of this kind of modelling for upper secondary education. The paper ends with implications for practice and research.

2. Theoretical background

2.1 Mathematical modelling

Mathematical modelling is a complex, iterative process, often depicted by a mathematical modelling cycle comprising a number of key stages. According to Galbraith and Stillman (2006), these stages are: A) Messy real-world problem, B) Real-world problem statement, C) Mathematical model, D) Mathematical solution, E) Real-world meaning of solution, F) Evaluation, and G) Report. In advancing through these stages, students should complete several learning activities involved in the transitions between them. These activities are: understanding, structuring, simplifying, interpreting context (transition A→B); assuming, formulating, mathematizing (B→C); working mathematically (C→D); interpreting mathematical output (D→E); comparing, critiquing, validating (E→F); communicating, justifying, report writing, if model is found appropriate (F→G); and revisiting the modelling process, if the model is found to be unsatisfactory (F→B). These researchers stressed that mathematical modelling often requires going back and forth among these stages. For example, going back and forth between stages Real world problem statement and Mathematical model (B→C), is critical in developing a mathematical model that captures the main features expressed in that statement.

There are other descriptions of a mathematical modelling cycle proposed in the literature, which also assume that mathematical modelling is a complex, non-linear process (e.g. Blum & Leiß, 2007; Niss 2015). Blum and Leiß (2007) used the following key Stages and main *activities* in transitions between them: Real situation – *understanding* – Situation model – *simplifying/structuring* – Real model – *mathematizing* – Mathematical model – *working mathematically* – Mathematical results – *interpreting* – Real results – *validating.* Niss (2015) used another framework. Its main components are: Preparing-making assumptions, choosing and formulating questions to answer; Mathematizing-developing a mathematical model by expressing features, relationships, assumptions, and questions in mathematical terms; Mathematized solution-using mathematical knowledge, skills and reasoning to solve the model; De-mathematizing-interpreting the mathematical outcomes of the previous stage in terms of the context and the questions that initiated the process of modelling; and Validating-assessing the quality and relevance of the model applied and the answers obtained.

Although the three conceptualizations presented above are somewhat different, they all support the position that the key activities in modelling cycles (aimed at descriptive modelling) may be summarized as: preparing for modelling, developing a model, solving the model, interpreting the solution, and validating the modelling. If the focus is on prescriptive mathematical modelling, an activity recommending changes (i.e. recommending actions to take to change the world modelled) could be added.

2.2 Data modelling

The term "data modelling" has been used to denote different processes in different areas. In software engineering, for example, data modelling is used to describe the process of creating a data model to describe what data will be stored in a database and how it will be organized (Simison & Witt, 2005). Such a usage holds true for computing education (e.g. Birnbaum & Langmead, 2017, p. 80). In statistics, on the other hand, this term denotes the search for and application of a stochastic data model that might generate the data in question (Breiman, 2001). The term data modelling has also been used in mathematics and statistics education (e.g. English, 2014; Kadijevich, 2016), and its conceptualization, as clarified below, would be viewed as an instance of mathematical modelling described above, in Section 2.1.

Data modelling in mathematics education is considered to be a developmental process, mostly studied by English (2012, 2014). She assumed that data modelling begins with students' inquiries into meaningful phenomena. After that, students identify various attributes of these phenomena. Then, by using these attributes, they organize, structure, visualize, and represent data (English, 2012). According to English and Sriraman (2010, p. 280), this modelling, as an instance of mathematical modelling, comprises the following six components: posing questions, generating and selecting attributes, measuring attributes, organizing data, displaying and representing data, and drawing inferences. In addition, they make clear that the developed model, which should enable students to answer their own initial questions, may be tested and revised, finally to be used to draw inferences and make recommendations. To improve data modelling achievement, English (2014) suggested, among other things, more practice in using multiple representations, modifying and combining representations, and selecting suitable representations. Due to modern technology, such a practice may, for example, be promoted by using interactive charts and dashboards, as will be exemplified in Section 3.1.

English's research on data modelling (e.g. English, 2012) has focussed on primary education, whilst the content of this paper deals with upper secondary education. However, the components of data modelling used in my research were similar to those described by English and Sriraman (2010). In my research, the components (I called them activities) were: Asking questions, Preparing data, Visualizing data, Answering questions, Validating modelling, and Recommending changes. As my data modelling also starts with the examination of meaningful phenomena (e.g. Females earn less than males!), it is clear that data modelling conceptualised in this way would be viewed as an instance of mathematical modelling. Indeed, as activities of asking questions and preparing data parallel that of preparing for modelling, the activity of visualizing data involves the activities of developing a model and solving it, and the activity of answering questions parallels that of interpreting a solution. In particular, there are many common sub-activities present in both data modelling and mathematical modelling. Bearing in mind Galbraith and Stillman (2006), these common sub-activities are: clarify problem context and questions, identify strategic entities (transition A→B); identify dependent and independent variables, specify calculations needed (B→C); use technology to perform calculations, and produce charts and tables (C→D); give meaning to mathematical results, select arguments to justify interpretations (D→E); and question model adequacy, examine implications of results (E→F).

Conceptualized in that way (with the activity of recommending changes), my approach to data modelling could be viewed as an instance of prescriptive modelling. Despite the recent demand to foster prescriptive modelling as well (e.g. Niss, 2015), the activity of recommending changes has not, to my knowledge, been examined as a separate component of a modelling cycle in relation to its other components (not only neighbouring ones). Apart from process cyclicity (e.g. Niss, 2015), we may focus on particular relations that may help us conceptualize a better distinction between descriptive and prescriptive data modelling (e.g. Are relations among preparing data, visualizing data, and validating modelling stronger in prescriptive modelling?).

As I will describe in Section 3, my approach to data modelling primarily makes use of interactive charts (descriptive, exploratory graphical displays), which are produced by visualizations based on simple mathematical models (e.g. frequencies, sums, and means). Implemented in this way, my data modelling is different from the application of statistical techniques to answer real-world problems. Statistics are concerned with particular stochastic data models, p values, and making inferences about the underlying population, whereas my data modelling is an instance of exploratory data analysis (Turkey, 1977), only aiming at discovering useful patterns, trends, effects, and interactions in the data examined.

3. Data modelling with interactive charts

In this section, I will first examine powerful, yet simple business intelligence tools, called interactive charts. After that I will examine data modelling with these charts, focusing on the main challenges in using them with respect to the key data modelling activities introduced in Section 2.2.

3.1 Interactive charts

Descriptive analyses that summarize data (e.g. by respective sums, averages, or counts) are frequently implemented in the form of interactive charts, which enable the viewing (and modelling) of multivariate phenomena in a dynamic way. Although created for the business world, these charts have entered into many fields, including computer-based learning (e.g. learning analytics).

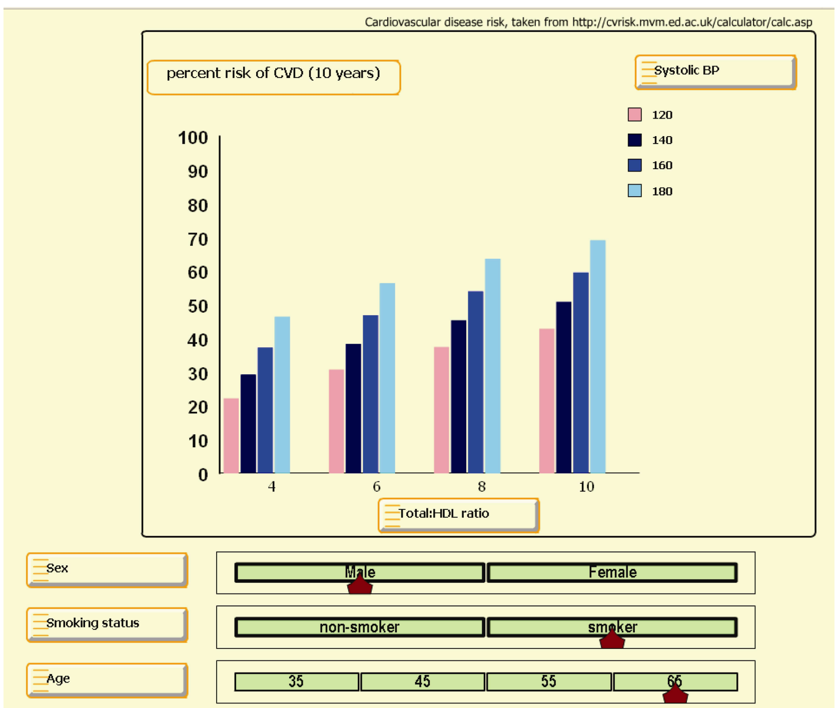

Fig. 1. An example of interactive chart: the risk of cardiovascular disease

Several interesting interactive charts may, for example, be found at `https://www.dur.ac.uk/smart.centre/`. A screenshot of one of these charts [the risk of cardiovascular disease] is presented in Figure 1. The modeller can consider several dimensions (in this case Sex, Smoking status, Age, and Systolic Blood Pressure (BP)), and use one of them as the legend field (e.g. Systolic BP). Any change (e.g. focus on female, or exchange the roles of Age and Systolic BP) updates the chart automatically.

Interactive charts are frequently implemented by using pivot tables and pivot charts. These pivot tools are available in a number of computing environments, such as the Microsoft Office suite, NumberGo (payware), and OpenOffice or LibreOffice (freeware). The pivot objects allow some summary parameters (e.g. the sum or the average of the values considered) to be shown in relation to different levels of detail. The user (modeller) selects these details with respect to his or her requirements, and, if necessary, collapses or breaks them down further (so-called roll-up or drill-down views, respectively). If, for example, the data analyzed refer to the sales of a number of items in several stores located in various cities, the pivot chart may display the total income by city, or, by going further with the analysis, the total income by city and store, or the total income by city, store, and item.

Although interactive charts can deal with several variables (e.g. one dependent and four independent, as in Figure 1), we can only examine one chart at a time. In other words, we cannot examine two charts side by side unless we have somehow made their copies previously. This limitation can be removed by using a dashboard, which is, in brief, a set of interactive reports, primarily interactive charts. A screenshot of a dashboard is given in Figure 2. As the value of its Region filter variable changes, the two charts and its Total summary measure update automatically.

Like interactive charts, dashboards are usually designed in a visual environment using the drag-and-drop approach. This environment may not only support the use of various types of charts and summary measures, but also enable the preparation of data by, for example, querying relevant datasets (e.g. ZOHO Reports at `https://www.zoho.com/reports/dashboard.html`). Data modelling using dashboards is examined in detail in Kadijevich (2016).

3.2 Data modelling

The key activities in my approach to data modelling, introduced in Section 2.2, were Asking questions, Preparing data, Visualizing data, Answering questions, Validating modelling, and Recommending changes. Apart from Visualizing data, the use of interactive charts would support other data modelling steps, especially Answering questions and Validating modelling. Through the former activity, modellers match patterns, trends, effects (and interactions) found with the questions posed. Through the latter activity, modellers improve the modelling applied by using other variables or charts, or even other data or another interactive charts tool.

If the students have not been given data sets to model, the use of interactive charts may also support (though not primarily) the step of Preparing data. This is because before interactive charts or dashboards can be built and used, data should be prepared, by applying database queries or different organizations of

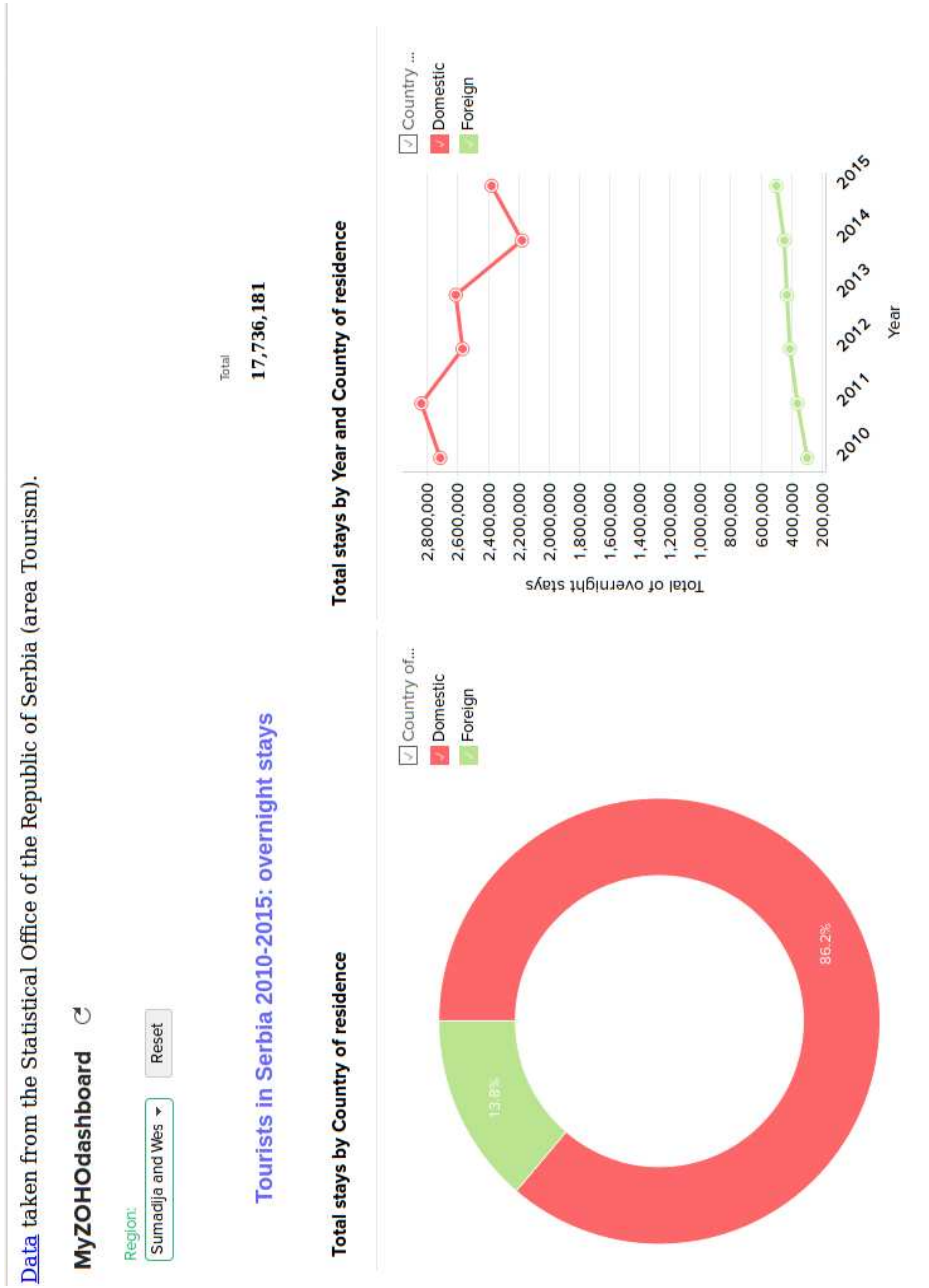

Fig. 2 An example of dashboard: Overnight stays by tourists in Serbia (available at `http://www.mi.sanu.ac.rs/~djkadij/Dashboard.htm`)

available data. Note that data scientists—those who apply various techniques from mathematics, statistics, and computer science to find useful patterns and trends in (big) data—spend most of their time collecting, cleaning, and organizing data (see Data Science Reports for 2016 and 2017 at `https://www.crowdflower.com`, for example).

3.2.1 Challenges. Challenges in using interactive charts relate to (1) data to use, (2) key data modelling activities, and (3) interactive charts tool to use (Kadijevich, 2016).

Several challenges may be met in the activity of preparing data, regarding various aspects of data preparation, such as querying data sources (Cox & Nikolopoulou, 1997), organizing available data in different ways (O'Donnell, 2005), and adding new variables, especially if these variables are based upon complex calculations. At an advanced level of preparing data (suitable for college level and above), challenges may be generated by data transformation (e.g. obtaining inflation-free normalizations as suggested by Tufte, 2001; addressing the issues of outliers and missing values), as well as data provenance and quality (e.g. data of poor quality, Sebastian-Coleman, 2013; lack of better indicators, Johnston et al., 2009; corruptibility of indicators, Ridgway et al., 2013).

As already mentioned, the activity of visualizing data in my approach to data modelling involves the classical mathematical modelling activities of developing a model and solving it. Developing a model, which basically requires modellers to identify dependent and independent variables and specify the calculations needed, may generate a number of challenges. These challenges may be the result of erroneous calculations, inappropriate selections of variables, confusion between dependent and independent variables, or using more, or fewer, independent variables than needed. On the other hand, solving the model, which basically requires modellers to perform calculations and produce charts and tables, is more or less challenge-free.

In the case of mathematical modelling in general, preparing domain knowledge for modelling and validating models and suggestions they support usually generate challenges (Niss, 2015). Firstly, in preparing for modelling (i.e. the activities of asking questions and preparing data in this approach to data modelling), the questions asked should match appropriate and available sets of variables that may be challenging. Secondly, in validating modelling, apart from the (great) difficulty of examining results in context (e.g. Gal & Trostianitser, 2016)—which is not related to interactive charts—challenges are generated by possible improvements in modelling based upon using other variables or charts, or even other data or another interactive charts tool.

Regarding challenges relating to which tool to use, it should primarily support appropriate presentations of multiple variables (adapted from Tufte, 2001). Bearing in mind Figure 2, if dashboards are used, they should support work with several variables by using various chart types and summary statistics. In addition, a dashboards builder should respect a number of design requirements (e.g. flexible solutions that are easy to upgrade; simple charts that improve perception; Yigitbasioglu & Velcu, 2013). Having all these in mind, it may indeed be a struggle to find a suitable tool, especially freeware (McNamara & Hansen, 2014).

4. Empirical study

4.1 Rationale

The complexity or difficulty of modelling tasks has occasionally been examined (e.g. examining task complexity in terms of linguistic, conceptual, mathematical, intellectual, representational, and contextual complexity, Stillman & Galbraith, 2003; examining the difficulty of modelling tasks in terms of a thought structure model, Reit & Ludwig, 2015). In addition, regarding the difficulty of modelling activities, research has evidenced that, for example, developing mathematical models (i.e. the transition from real world problem statement to mathematical model) is one of the most difficult modelling activities (Galbraith & Stillman, 2006). However, the difficulty of key modelling activities has not been compared. Since the results of this comparison would be different for different modelling tasks, the examination only makes sense if done for a particular class of modelling tasks, whose results would improve both research (e.g. understanding the main challenges and reasons for them) and practice (e.g. developing appropriate scaffolds that support students' modelling). This study thus examined and compared the difficulty of key data modelling activities by using an example of one class of modelling tasks solved by novice students. Being aware of a number of recent demands to include data modelling in secondary education, this study also examined the importance of data modelling with interactive charts for upper secondary education.

4.2 Research questions

The study used the following research questions: (1) to what extent might a particular data modelling activity be difficult for students to complete successfully; and (2) to what extent might data modelling with interactive charts be important for upper secondary education?

These research questions were answered for data modelling with commonly used interactive charts in which a file with data to consider (comprising 5–6 variables) and a short description of the underlying context were given to students. My decision to simplify modelling in such a way was influenced by Galbraith and Stillman (2006). To attain better mathematical modelling, these researchers wanted to keep mathematical, technological, and other prerequisites as simple as possible.

In answering these questions, I used an international sample of technology-skilled participants. Such a sample was used not only to avoid the possible negative impact of participants' insufficient technological competences in their responses, but also to consider responses from different educational systems, which would result in stronger findings.

4.3 Method

4.3.1 Sample. The research participants were all CASIO experts/educators who had attended the Pan-European Educational Conference "Exploring Mathematics with Technology", Budapest-Hungary, 13–15 October 2017. This group meets annually to discuss opportunities and challenges of using digital mathematics tools in classrooms. In November 2017, all participants at this conference (around forty from fifteen European countries) were invited, via e-mail, to take part in an online survey. Nineteen educators, from more than ten countries across Europe, completed this survey, providing the sample for my empirical study. Eleven of these nineteen participants (58%) were teachers, eight (42%) were females, and eight (42%) had more than twenty hours experience of data modelling with interactive charts.

4.3.2 Variables and design. Three background variables were used: *gender* (with values 1-female, and 2-male), *experience* (values: 1-0-20 hours, 2-20+ hours), and *occupation* (values: 1-teacher, 2-researcher, designer, policy maker, or industry partner).

The importance of data modelling with interactive charts for upper secondary education (variable *importance*) was estimated on a 5-point scale. The five points were: 1-totally unimportant, 2-not important, 3-neutral, 4-important, and 5-very important.

For each modelling activity, its difficulty for students to complete successfully was also estimated on a 5-point scale. These five points were: 1-very easy, 2-easy, 3-moderate, 4-difficult, and 5-very difficult. There were five modelling variables in total, because the activity of preparing data was not examined. It was assumed that a file with data to consider had been given to students.

Since this study primarily dealt with the means of related samples, it applied a repeated measures design. The difficulty of modelling activity (considered as a dependent variable *difficulty*) was examined in terms of these five modelling variables (taken as the values of a factor *activity*). The statistical analysis used was a repeated measures analysis of variance (ANOVA).

4.3.3 Instrument and procedures. This study used an on-line survey comprising five questions regarding the nine variables mentioned above. Content of this survey is given at `http://www.mi.sanu.ac.rs/~djkadij/DMS.pdf`. The survey was developed and administered by the author of this paper, who used a free version of the SurveyMonkey environment for this purpose. The two main questions were:

1. For each activity below, please estimate to what extent it would be difficult for students to complete successfully.

	Very easy	Easy	Moderate	Difficult	Very difficult
Asking questions	○	○	○	○	○
Visualizing data	○	○	○	○	○
Answering questions	○	○	○	○	○
Validating modelling	○	○	○	○	○
Recommending changes	○	○	○	○	○

2. To what extent did you find data modelling with interactive charts important for upper secondary education?

○ Totally unimportant ○ Not important ○ Neutral ○ Important ○ Very important

As already mentioned, all participants at the CASIO 2017 Budapest conference were invited to take part in this survey in November 2017 via e-mail. About half of them did so within ten days.[1]

The context for this survey was set up during the conference. After his presentation on e-learning, which exemplified data modelling with interactive charts, the author of this paper administered a survey that asked the conference participants to indicate anonymously one or two activities in data modelling with interactive charts at upper secondary level that would be, in general, the most difficult for novice students. Although, as mentioned above, findings in general may be of little use, one outcome is still noteworthy: two activities that rely more on technology use than on conceptual and contextual clarifications (Preparing data and Visualizing data) were, on average, indicated as the most difficult less often than the remaining four modelling activities (Asking questions, Answering questions, Validating modelling, and Recommending changes), which rely more on conceptual and contextual clarifications than on technology use. The relative average frequencies of selecting these two groups of activities were 13% and 37%, respectively ($N = 35$, $z = -2.135$, $p = 0.033$).

4.3.4 Data transformation and statistical analysis. To attain a more precise, interval measurement of the five difficulty variables, the participants' responses, previously expressed by corresponding numbers, were transformed into Guttman's (1955) image form scores (as, for example, done in Kadijevich (2006), and Kadijevich et al. (2016)). Because the reliability (Cronbach's alpha) of a summary variable regarding such a composed 5-item questionnaire was above 0.90 and the correlation of each item with this summary variable was also above 0.90, the reliability of each of the five Guttmanized difficulty variables was high (the formula for correction for attenuation, Wanous & Hudy, 2001, was applied).

Apart from basic descriptive statistics summarizing relative frequencies and means for the variables applied, two parametric and one non-parametric statistical tests were used. Firstly, I used a repeated measures ANOVA to test whether the mean of the estimated difficulty of particular activity was the same across the five activities considered, followed by a t-test for paired-samples, comparing all pairs of means. Secondly, for variable importance, I used a sign test to test whether its median was equal to 3 (i.e. scale point "neutral").

Before applying repeated measures ANOVA, the required sample size and assumptions for this analysis were checked.

- Regarding the sample size required for a repeated measures ANOVA, one recommendation calls for at least $10+k$ subjects, where k is the number of groups

[1] This is an acceptable response rate, recalling that average response rate in online surveys is around 30% (Saldivar, 2012).

or factor levels.[2] This was attained because the sample used in this study comprised 19 participants ($19 > 10 + 5$).

- There are five assumptions for this analysis[3], of which four were attained: use a continuous dependent variable (done); examine at least two related groups (there were five types of modelling activities); have no significant outliers present (attained according to Grubb's test with 2.681 as critical z value); and have the dependent variable in these groups approximately normally distributed (attained according to Shapiro-Wilk test). As the fifth assumption of sphericity was violated, the degrees of freedom (*df*) had to be corrected.

4.4 Results

As already mentioned, eleven participants (58%) were teachers, eight (42%) were females, and eight (42%) had more than twenty hours experience of data modelling with interactive charts. There were no statistically significant relationships among background variables gender, occupation, and experience.

The means (and standard deviations) of estimated difficulty of the five modelling activities were the following: 2.48 (0.73) – Asking questions; 2.30 (0.60) – Visualizing data; 2.64 (0.75) – Answering questions; 3.32 (0.73) – Validating modelling; and 3.48 (0.89) – Recommending changes. A repeated measures ANOVA (with a Greenhouse-Geisser *df* correction of 0.712)[4], which evaluated differences among these means, was significant ($F(2.848, 51.260) = 60.956$, $p = 0.000$; effect size: 0.772 – partial eta squared). Follow-up pairwise comparisons of means, done by a t-test for paired samples with Bonferroni correction, revealed that there were seven significant differences, all being significant at a 0.01 level. These seven differences were:

- Asking questions was estimated easier than Validating modelling and Recommending changes;
- Visualizing data was estimated easier than Answering questions, Validating modelling, and Recommending changes;
- Answering questions was estimated easier than Validating modelling and Recommending changes.

The median of variable importance was 4 (scale point "important"), and according to an exact sign test, it was significantly greater than 3 (scale point "neutral"), $p = 0.001$.

[2]See `http://www.real-statistics.com/anova-repeated-measures/sphericity/` for this recommendation. Simulations evidenced that a 19-subject sample is also appropriate for eight factor levels, provided that the assumption of normality is not violated (Oberfeld & Franke, 2013; Figure 4).

[3]For these assumptions, see, for example, `https://statistics.laerd.com/spss-tutorials/one-way-anova-repeated-measures-using-spss-statistics.php`

[4]The assumption of sphericity was violated: Mauchly's Test: 0.206, Chi-square = 25.916, $df = 9$, $p = 0.002$; Epsilon: 0.712 – Greenhouse-Geisser, 0.860 – Huynh-Feldt.

4.5 Discussion

Two important findings emerged from this empirical study. Firstly, on average, the estimated difficulty of modelling activity varied across modelling activities, which provided evidence that some modelling activities would be easier than others. Secondly, data modelling with interactive charts was found important for upper secondary education. The following sub-sections examine these findings in detail by referring to the two research questions used in this study.

4.5.1 First question. The first research question examined the extent to which a particular data modelling transitional activity would be difficult for students to complete successfully. The estimated difficulty, on average, ranged from 2.30 (Visualizing data) to 2.64 (Answering questions) to 3.48 (Recommending changes). Because of the six differences in pair of means found (any of the first three activities was estimated easier than any of the remaining two), I additionally compared means of two groups of activities, namely: Activities performed (Asking questions, Visualizing data, and Answering questions), and Activities neglected (Validating modelling and Recommending changes). These group labels were chosen to reflect students' common practice in traditional modelling, where validating modelling is usually missing (e.g. Blum & Borromeo, 2009; Blum 2015). Even when students are explicitly required to validate models developed and to suggest actions based on their outcomes, most of them would rather deal with other modelling activities (e.g. Kadijevich, 2012). By using average values of participants' Guttmanized scores regarding the two groups of activities, it was found that Activities performed were, on average, estimated by technology-skilled educators as easier than Activities neglected (Activities performed: M = 2.48, SD = 0.66; Activities neglected: M = 3.40, SD = 0.79; $t = 13.534$, $df = 18$, $p = 0.000$).

This finding was not surprising, having in mind the main challenges in data modelling activities discussed in Section 3.2.1. Although both groups of activities are supported by providing students with data and contextual information to consider, contrary to Activities performed, Activities neglected rely more on conceptual and contextual clarifications than on technology use (probably to a similar extent as the estimated difficulty of Validating modelling (3.32) was equal statistically to that of Recommending changes (3.48)). These clarifications are very challenging (e.g. English, 2012; Niss, 2015) because the results should be examined in context, possibly asking to improve modelling by using, in my approach, other variables or charts, or even other data or another interactive charts tool. Completing these tasks successfully is usually out of the reach of most students, most probably due to (1) limited stills in problem structuring (Kadijevich, 2012), and (2) insufficient knowledge of context under scrutiny (generalized from Ben-Zvi, 2002), which may interact with subject and technology domains in a complex way (possibly resembling the structure of tapestry, Stillman, 1998).

4.5.2 Second question. The second research question examined the extent to which data modelling with interactive charts was perceived as important for upper secondary education by technology-skilled educators. Regarding the scale points applied, one participant chose “not important”, four chose “neutral”, whilst the remaining fourteen chose “important” (8) or “very important” (6). According to the sign test applied, the median of these responses (4 i.e. “important”) was statistically significantly greater than the middle scale point (3 or “neutral”). (The number of those who chose “important” or “very important” was statistically significantly greater than the number of others who chose “neutral”, “not important” or “totally unimportant”; $\chi 2 = 4.263$, $df = 1$, $p = 0.039$.) In other words, the participants did find data modelling with interactive charts important for upper secondary education. Although there have been a number of demands recently for the inclusion of data modelling in secondary education (e.g. Davison, 2015; Engel, 2017; Gould et al., 2016; Metz, 2015; Ridgway, 2016), no empirical data on the importance of this inclusion have been used to support these calls, according to my reading.

The reliability of the single-item measure of variable importance should not be questioned because even a complex (possibly multidimensional) construct like mathematical anxiety can be measured reliably in that way (Nez-Pea et al., 2014). If the meaning of an object in such a measure does not call for the consideration of its concrete components, this measure should be accepted (Bergkvist & Rossiter, 2007). Note that there were no statistically significant relationships between importance and any of the remaining variables used in this empirical study, which provided evidence of suitability of the single-item measure of variable importance.

5. Conclusion

This paper examined data modelling using interactive charts in upper secondary education. The examination showed that this kind of modelling could be viewed as an instance of prescriptive mathematical modelling. By referring to key modelling activities, it also summarized the main challenges in using these charts, underlying considerable difficulties in activities based upon conceptual and contextual clarifications. Although as pilot study used a small sample of technology-skilled educators, it confirmed these difficulties because these educators estimated that Activities performed (Asking questions, Visualizing data, and Answering questions) were, on average, easier than Activities neglected (Validating modelling and Recommending changes, which were equally difficult, statistically). As already mentioned, these group labels were chosen to reflect students’ common practice in traditional modelling (e.g. Blum, 2015; Kadijevich, 2012). This study also supported the belief that data modelling with interactive charts is perceived by these educators to be important for upper secondary education.

Because of difficulties with conceptual and contextual clarifications, it is always important to base modelling on contexts familiar to students (Henry Pollak, personal communication, July 26, 2016). However, students may not be (that) familiar even with contexts they choose themselves (Kadijevich, 2012). Because contextual knowledge and problem structuring skills probably influence each other (suggested by Restrepo & Christiaans, 2004), we may help data modellers develop problem structuring skills (e.g. selecting variables, measures, table, and charts to use) while improving knowledge of context under scrutiny (e.g. clarifying what other issues to examine, how they have been measured, and for what points in time to use data available), and vice versa.

The study examined an example of a particular class of modelling tasks: data modelling with commonly used interactive charts in which data and contextual information to consider have been given to students. Such a didactical approach to modelling requires the omission of a particular modelling activity (e.g. Preparing data as also done in Ridgway, 2016), which would reduce the difficulties typically found in other modelling activity (e.g. starting from Asking questions may be quite difficult, Gould et al., 2016). (Data cleaning is particularly challenging [Han, Kamber & Pei, 2012] and should be avoided or postponed as there are several ways to address particular data oddities.) This didactical approach may also calls for assessing the outcomes of particular modelling activity undertaken by other modellers (cf. for asking questions about data-based statements, Schiller & Engel, 2016), which would contribute to a better understanding of this activity and other activities examined in this assessment (e.g. Visualizing data to check these data-based statements). As there is slim evidence of how data modelling with interactive charts would be taught (some didactical remarks are given in Kadijevich, 2016), the implementation of this approach could improve matters. In doing so, the difficulties of particular modelling activities may be captured and analyzed in terms of a suitable framework. Having in mind Stillman and Galbraith (2003), this framework may deal with conceptual, contextual, technological, mathematical, representational, and other sorts of difficulty. However, by applying an analogy with a requirement to use interactive charts with "a good balance between visual complexity and information utility" (Yigitbasioglu & Velcu, 2013; p. 46) we should require to use a framework with a good balance between its complexity and pedagogical utility.

Although the research summarized in this paper basically dealt with the difficulties of activities in data modelling using interactive charts when some activities have been made easier for students with material given to them, let me close its content with a note on scaffolds, i.e. hints and supports that would enable students to complete, on their own, modelling activities successfully (these aids in modelling are examined in Kaiser & Stender, 2013, for example). Bearing in mind that technology integration depends on making connections among content, pedagogy and technology (Niess & Gillow-Wiles, 2017), instruction may primarily apply scaffolds that make links between contextual/conceptual and technology-related issues (e.g. between questions to ask and chart types to use, visualizations produced and questions to answer, or modelling features to validate and technology components used). I hope that the development and application of such scaffolds would help teachers and students connect key activities in data modelling with technology and attain better results of this modelling.

Acknowledgements. This contribution resulted from the author's work on the project "Improving the quality and accessibility of education in modernization processes in Serbia" (No. 47008), financially supported by the Ministry of Education, Science, and Technological Development, Republic of Serbia (2011–2018). The author dedicates the contribution to his son Aleksandar.

REFERENCES

[1] Ben-Zvi, D. (2002). *Seventh grade students' sense making of data and data representations*. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics (CD-ROM)*. Voorburg, The Netherlands: International Statistical Institute.

[2] Bergkvist, L., & Rossiter, J. R. (2007). *The predictive validity of multiple-item versus single-item measures of the same constructs*. Journal of Marketing Research, 44 (2), 175–184.

[3] Birnbaum, D. J., & Langmead, A. (2017). *Task-driven programming pedagogy in the digital humanities*. In S. Fee, A. Holland-Minkley, & T. Lombardi (Eds.), *New directions for computing education* (pp. 63–85). Cham, Switzerland: Springer.

[4] Blum, W. (2015). it Quality teaching of mathematical modelling: What do we know, what can we do?, In S. J. Cho (Ed.), *The Proceedings of the 12th International Congress on Mathematical Education* (pp. 73–96). Cham, Switzerland: Springer.

[5] Blum, W., & Borromeo Ferri, R. (2009). *Mathematical modelling: Can it be taught and learnt?*, Journal of Mathematical Modelling and Application, **1** (1), 45–58.

[6] Blum, W., & Leiß, D. (2007). *How do students and teachers deal with modelling problems*. In C. Haines, P. Galbraith, W. Blum, & S. Khan (Eds.), *Mathematical modelling: Education, engineering, and economics* (pp. 222–231). Chichester, UK: Horwood.

[7] Breiman, L. (2001). *Statistical modelling: The two cultures*. Statistical Science, **16** (3), 199–231.

[8] Cox, M., & Nikolopoulou, K. (1997). *What information handling skills are promoted by the use of data analysis software?*, Education and Information Technologies, **2** (2), 105–120.

[9] Davis, P. J., & Hersh, R. (1990). *Descartes' Dream*. London: Penguin.

[10] Davison, C. B. (2015). *Addressing the challenges of teaching big data in technical education*. The CTE Journal, **3** (1), 43–50.

[11] Engel, J. (2017). *Statistical literacy for active citizenship: A call for data science education*. Statistics Education Research Journal, **16** (1), 44–49.

[12] English, L. D. (2012). *Data modelling with first-grade students*. Educational Studies in Mathematics, **81**, 15–30.

[13] English, L. D. (2014). *Promoting statistical literacy through data modelling in the early school years*. In E. J. Chernoff, & B. Sriraman (Eds.), *Probabilistic thinking: Presenting plural perspectives* (pp. 441–457). Dordrecht, the Netherlands: Springer.

[14] English, L. D. (2016). *STEM education K-12: Perspectives on integration*. International Journal of STEM Education, **3** (3), 1–8.

[15] English, L., & Sriraman, B. (2010). *Problem solving for the 21st century*. In B. Sriraman, & L. English (Eds.), *Theories of mathematics education. Advances in mathematics education* (pp. 263–290). Berlin: Springer.

[16] Gal, I., & Trostianitser, A. (2016). *Understanding basic demographic trends: Connecting table reading, task design, and context*. In J. Engel (Ed.), *Promoting understanding of statistics about society*. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE.

[17] Galbraith, P. L, & Stillman, G. (2006). *A framework for identifying student blockages during transitions in the modelling process*. Zentralblatt für Didaktik der Mathematik, **38** (2), 143–162.

[18] Gould, R. et al. (2016). *Teaching data science to secondary students: The mobilize introduction to data science curriculum*. In J. Engel (Ed.), *Promoting understanding of statistics about society*. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE.

[19] Greca, I. M., & Moreira, M. A. (2000). *Mental models, conceptual models, and modeling.* International Journal of Science Education, **22** (1), 1–11.

[20] Guttman, L. (1955). *Image theory for the structure of quantitative variates.* Psychometrika, **18** (4), 277–296.

[21] Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed). Amsterdam: Elsevier.

[22] Johnston, R. B. et al. (2009). *Addressing information gaps (IMF staff position note).* Washington, DC: International Monetary Fund.

[23] Jonassen, D. H. (2006). *Modeling with Technology: Mindtools for Conceptual Change* (3rd ed). Upper Saddle River, NJ: Prentice Hall.

[24] Kadijevich, Dj. (2006). *Developing trustworthy TIMSS background measures: A case study on mathematics attitude.* The Teaching of Mathematics, **9** (2), 41–51.

[25] Kadijevich, Dj. M. (2012). *Examining errors in simple spreadsheet modelling from different research perspectives.* Journal of Educational Computing Research, **47** (2), 137–153.

[26] Kadijevich, Dj. M. (2016). *Data modelling with dashbards: Opportunities and challenges.* In J. Engel (Ed.), *Promoting understanding of statistics about society.* Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE.

[27] Kadijevich, Dj. M., Odovic, G., & Maslikovic, D. (2016). *Using ICT and quality of life: Comparing persons with and without disabilities.* In K. Miesenberger, C. Bühler, & P. Penaz (Eds.), *Computers helping people with special needs.* ICCHP 2016. Lecture Notes in Computer Science, 9758, 129–133.

[28] Kaiser, G., & Stender, P. (2013). *Complex modelling problems in co-operative, self-directed learning environments.* In G. Stillman, G. Kaiser, W. Blum, & J. Brown (Eds.), *Teaching mathematical modelling: Connecting to research and practice* (pp. 277–293). Dordrecht, the Netherlands: Springer.

[29] Metz, S. (2015). *Editor's corner: Big data.* The Science Teacher, **82** (5), 6.

[30] McNamara, A., & Hansen, M. (2014). *Teaching data science to teenagers.* In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the Ninth International Conference on Teaching Statistics* (CD-ROM). Voorburg, the Netherlands: International Statistical Institute.

[31] Niess, M. L., & Gillow-Wiles, H. (2017). *Expanding teachers' technological pedagogical reasoning with a systems pedagogical approach.* Australasian Journal of Educational Technology, **33** (3), 77–95.

[32] Niss, M. (2015). *Prescriptive modelling – challenges and opportunities.* In G. A. Stillman, W. Blum, & M. S. Biembengut (Eds.), *Mathematical modelling in education research and practice: Cultural, social and cognitive influences* (pp. 67–79). Cham, Switzerland: Springer.

[33] Núñez-Peña, M. I., Guilera, G., & Suárez-Pellicioni, M. (2014). *The single-item math anxiety scale (SIMA): An alternative way of measuring mathematical anxiety.* Journal of Psychoeducational Assessment, **32** (4) 306–317.

[34] Oberfeld, D., & Franke, T. (2013). *Evaluating the robustness of repeated measures analyses: The case of small sample sizes and nonnormal data.* Behavior Research Methods, **45** (3), 792–812.

[35] O'Donnell, P. (2005). *The problem with pivot tables.* Business Intelligence Journal, **10** (1), 25–31.

[36] Reit, X.-R., & Ludwig, M. (2015). *An approach to theory based modelling tasks.* In G. A. Stillman, W. Blum, & M. S. Biembengut (Eds.), *Mathematical modelling in education research and practice: Cultural, social and cognitive influences* (pp. 81–91). Cham, Switzerland: Springer.

[37] Restrepo, J., & Christiaans, H. (2004). *Problem structuring and information access in design.* Journal of Design Research, **4** (2).

[38] Ridgway, J. (2016). *Implications of the data revolution for statistics education.* International Statistical Review, **84** (3), 528–549.

[39] Ridgway, J., James, N., & McCusker, S. (2013). *Open data' and the semantic web require a rethink on statistics teaching.* Technology Innovations in Statistics Education, **7** (2).

[40] Saldivar, M. G. (2012). *A primer on survey response rate*. Tallahassee, FL: Learning Systems Institute, Florida State University.

[41] Schiller, A., & Engel, J. (2016). *Civic statistics and the preparation of future secondary school mathematics teachers*. In J. Engel (Ed.), *Promoting understanding of statistics about society*. Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. The Haag, the Netherlands: ISI/IASE.

[42] Sebastian-Coleman, L. (2013). *Measuring data quality for ongoing improvement: A data quality assessment framework*. Amsterdam: Elsevier.

[43] Simison, G. C., & Witt, G. C. (2005). *Data Modelling Essentials* (3rd ed). Amsterdam: Morgan Kaufmann.

[44] Stillman, G. (1998). *The emperor's new clothes? Teaching and assessment of mathematical applications at the senior level*. In P. Galbraith, W. Blum, G. Booker, G., & D. Huntley (Eds.), *Mathematical modelling: Teaching and assessment in a technology-rich world* (pp. 243–253). Chichester, UK: Horwood.

[45] Stillman, G., & Galbraith, P. (2003). *Towards constructing a measure of the complexity of application tasks*. In S. J. Lamon, W. A. Parker, & S. K. Houston (Eds.), *Mathematical modelling: A way of life–ICTMA 11* (pp. 179–188). Chichester, UK Horwood.

[46] Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.

[47] Tukey, J. W. (1977). *Exploratory Data Analysis*. Reading, PA: Addison-Wesley.

[48] Wanous, J. P., & Hudy, M. J. (2001). *Single-item reliability: A replication and extension*. Organizational Research Methods, **4** (4), 361–375.

[49] Yigitbasioglu, O. M., & Velcu, O. (2013). *A review of dashboards in performance management: Implications for design and research*. International Journal of Accounting Information Systems, **13** (1), 41–59.

Institute for Educational Research, Dobrinjska 11/III, 11000 Belgrade, P.O.B. 546, Serbia

E-mail: `djkadijevic@ipi.ac.rs`