Liu and Huang *Smart Learning Environments* (2017) 4:1
DOI 10.1186/s40561-016-0040-4

Smart Learning Environments

REVIEW Open Access

# The use of data science for education: The case of social-emotional learning

Ming-Chi Liu and Yueh-Min Huang*

* Correspondence:
huang@mail.ncku.edu.tw
Department of Engineering Science,
National Cheng Kung University,
No. 1, University Road, Tainan
70101, Taiwan

## Abstract

The broad availability of educational data has led to an interest in analyzing useful knowledge to inform policy and practice with regard to education. A data science research methodology is becoming even more important in an educational context. More specifically, this field urgently requires more studies, especially related to outcome measurement and prediction and linking these to specific interventions. Consequently, the purpose of this paper is first to incorporate an appropriate data-analytic thinking framework for pursuing such goals. The well-defined model presented in this work can help ensure the quality of results, contribute to a better understanding of the techniques behind the model, and lead to faster, more reliable, and more manageable knowledge discovery. Second, a case study of social-emotional learning is presented. We hope the issues we have highlighted in this paper help stimulate further research and practice in the use of data science for education.

**Keywords:** Data science, Social-emotional learning, Education

## Introduction

Recently, AlphaGo, an artificially intelligent (AI) computer system built by Google, was able to beat world champion Lee Sedol at a complex strategy game called Go. AlphaGo's victory shocked not only artificial intelligence experts, who thought such an event was 10 to 15 years away, but also educators, who worried that today's high-value human skills will rapidly be sidelined by advancing technology, possibly even by 2020 (World Economic Forum 2016). Such potential technologies also catch some reflections of the relevance of certain educational practices in the future.

At the same time, emerging AI technologies not only pose threats but also create opportunities of producing a wide variety of data types from human interactions with these platforms. The broad availability of data has led to increasing interest in methods for exploring useful knowledge relevant to education—the realm of data science (Heckman and Kautz 2013; Levin 2013; Moore et al. 2015). In other words, data-driven decision-making through the collection and analysis of educational data is increasingly used to inform policy and practice, and this trend is only likely to grow in the future (Ghazarian and Kwon 2015).

The literature on education data analytics has many materials on the assessment and prediction of students' academic performance, as measured by standardized tests (Fernández et al. 2014; Linan and Perez 2015; Papamitsiou and Economides 2014; Romero and Ventura 2010). However, research on education data analytics should go beyond explaining student success with the typical three Rs (reading, writing and arithmetic) of literacy in the current economy (Lipnevich and Roberts 2012). Furthermore, the availability of data alone does not ensure successful data-driven decision-making (Provost and Fawcett 2013). Consequently, there is an urgent need for further research on the use of an appropriate data-analytic thinking framework for education. The purpose of this paper is first to identify research goals to incorporate an appropriate data-analytic thinking framework for pursuing such goals, and second to present a case study of social-emotional learning in which we used the data science research methodology.

## Defining data science

Dhar (2013) defines data science as the study of the generalizable extraction of knowledge from data. At a high level, Provost and Fawcett (2013) defines data science as a set of fundamental principles that support and guide the principled extraction of information and knowledge from data. Furthermore, Wikipedia defines data science (DS) as extracting useful knowledge from data by employing techniques and theories drawn from many fields within the broad areas of mathematics, statistics, and information technology. The field of statistics is the core building block of DS theory and practice, and many of the techniques for extracting knowledge from data have their roots in this. Traditional statistical analytics mainly have mathematical foundations (Cobb 2015); while DS analytics emphasize the computational aspects of pragmatically carrying out data analysis, including acquisition, management, and analysis of a wide variety of data (Hardin et al. 2015). More importantly, DS analytics follow frameworks for organizing data-analytic thinking (Baumer 2015; Provost and Fawcett 2013).

## Vision for future education

Character. Disposition. Grit. Growth mindset. Non-cognitive skills. Soft skills. Social and emotional learning. People use these words and phrases to describe skills that they also often refer to as nonacademic skills (Kamenetz 2015; Moore et al. 2015). Among these various terms, the social-emotional skills promoted by the Collaborative for Academic, Social and Emotional Learning (http://www.casel.org/) have mostly been accepted by the broader educational community (Brackett et al. 2012). A growing number of studies show that these nonacademic factors play an important role in shaping student achievement, workplace readiness, and adult well-being (Child Trends 2014). For example, Mendez (2015) finds that nonacademic factors play a prominent role in explaining variation in 15-years-old school children's' scholastic performance, as measured by the Program for International Students Assessment (PISA) achievement tests. Lindqvist and Vestman (2011) also find strong evidence that men who fare poorly in the labor market—in the sense of unemployment or low annual earnings—lack non-cognitive rather than cognitive abilities. Furthermore, Moffitt et al. (2011) find that the emotional skill of self-control in childhood is associated with better physical health, less substance dependence, better personal finances, and fewer instances of criminal offending in adulthood.

Due to a new understanding of the impact of nonacademic factors in the global economy, a growing movement in education has raised the focus on building social-emotional competencies in national curricula. In fact, countries like China, Finland, Israel, Korea, Singapore, the United States, and the United Kingdom currently mandate that a range of social-emotional skills be part of the standard curriculum (Lipnevich and Roberts 2012;

Ren 2015; Sparks 2016). The movement involves some complex issues ranging from the establishment of social and emotional learning standards to the development of social and emotional learning programs for students, and to the offering of professional development programs for teachers, and to the carrying out of social and emotional learning assessments (Kamenetz 2015).

However, as argued by Sparks (2016), research studying these skills has not quite caught up with their growing popularity. A number of authors raise various directions for future research in social and emotional learning. Child Trends (2014), for instance, conducted a systematic literature review of different social-emotional skills and highlighted the need for further research on the importance of the following five skills: self-control, persistence, mastery orientation, academic self-efficacy, and social competence. Moreover, Moore et al. (2015) provide conceptual and empirical justification for the inclusion of nonacademic outcome measures in longitudinal education surveys to avoid omitted variable bias, inform the development of new intervention strategies, and support mediating and moderating analyses. Likewise, Levin (2013) and Sellar (2015) both suggest that the development of data infrastructure in education should select a few nonacademic skill measures in conjunction with the standard academic performance measures. Furthermore, Duckworth and Yeager (2015) note that how multidimensional data on personal qualities can inform action in educational practice is another topic that will be increasingly important in this context.

Although all those issues have varying significances regarding the measurement and development of social and emotional learning, the following two research goals are priorities for studies of social and emotional learning:

1. Developing assessment techniques,
2. Providing intervention approaches.

These two research areas strongly affect the development of social-emotional skills, which are the principal concerns of the domains of education and data science, and which can be studied to derive evidence-based policies. To consider these issues, this paper focuses on (a) the suggested data science research methodology that is applicable to reach these goals, and (b) the case study of social-emotional learning in which we used the data science research methodology.

## Methodology review for data science

To better pursue those goals, it could be useful to formalize the knowledge discovery processes within a standardized framework in DS. There are several objectives to keep in mind when applying a systemic approach (Cios et al. 2007): (1) help ensure that the quality of results can contribute to solving the user's problems; (2) a well-defined DS model should have logical, well-thought-out substeps that can be presented to decision-makers who may have difficulty understanding the techniques behind the model; (3) standardization of the DS model would reduce the amount of extensive background knowledge required for DS, thereby leading directly to a knowledge discovery process that is faster, more reliable, and more manageable.

In the context of DS, the Cross-Industry Standard Process for Data Mining (CRISP-DM) model is the most widely used methodology for knowledge discovery (Guruler and

Istanbullu 2014; Linan and Perez 2015; Shearer 2000). It has also been incorporated into commercial knowledge discovery systems, such as SPSS Modeler. To meet the needs of the academic research community, Cios et al. (2007) further develop a process model based on the CRISP-DM model by providing a more general, research-oriented description of the steps. Applications of Cios et al. process model follow six steps, as shown in Fig. 1.

### Understanding of the problem domain

This initial step involves thinking carefully about the use scenario, understanding the problem to be solved and determining the research goals. Working closely with educational experts helps define the fundamental problems. Research goals are structured into one or more DS subtasks, and thus, the initial selection of the DS tools (e.g., classification and estimation) can be performed in the later step of the process. Finally, a description of the problem domain is generated.

An example research goal would be: Since meaningful learning requires motivation to learn, researchers are interested in real-time modeling of students' motivational orientations (e.g., approach vs. avoidance). Similarly, researchers might be interested in developing models that can automatically detect affective states (e.g., anxiety, frustration, boredom) from machine-readable signals (Huang et al. In Press; Lai et al. 2016; Liu et al. 2015).

### Understanding of the data

This step includes collecting sample data that are available and deciding which data, including format and size, will be needed. To better understand the strengths and limitations of the data, it also includes checking data completeness, redundancy, missing values, the plausibility of attribute values. Background knowledge can be used to guide these checks. Another critical part of this step is estimating the costs and benefits of each data source and deciding whether further investment in collection is worthwhile.

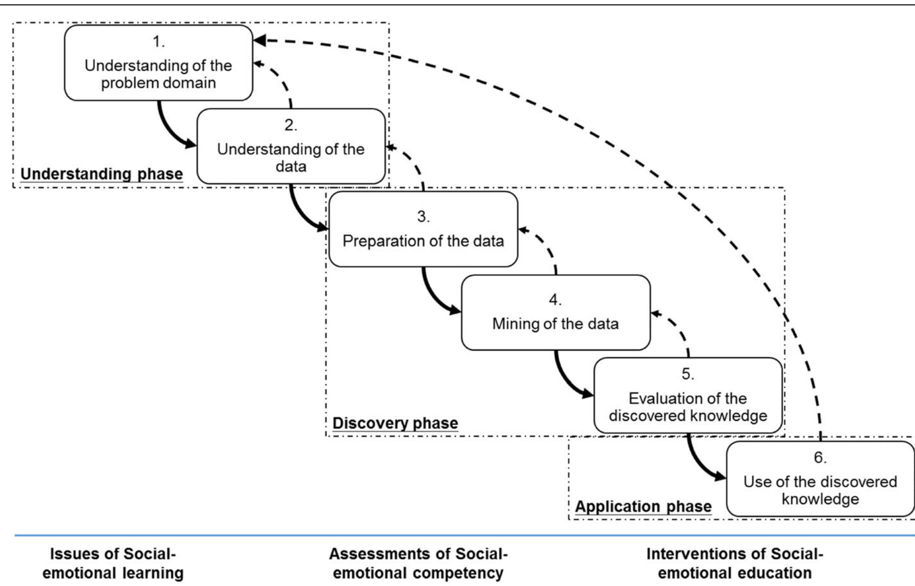


**Fig. 1** Cios et al.'s process model. Source: adapted from Cios and Kurgan (2005)

Finally, this step includes verifying that the data matches one or more DS subtasks in the last step.

For example, researchers may decide to analyze log traces in an online learning session to make inferences about students' motivational orientations. Moreover, researchers may choose to collect physiological data (such as facial expression, blood volume pulse, and skin conductance data) to develop models that can automatically detect affective states.

To date, DS has relied heavily on two data sources (Siemens 2013): student information systems (SIS, for in generating learner profiles, such as grade point averages) and learning management systems (LMS). For example, Moodle (https://moodle.org/) and Blackboard (http://www.blackboard.com/) can record logs for user activity in courses, forums, and groups. Linan and Perez (2015) suggest using Google Analytics to gather information about a site, such as the number of visits, pages visited, the average duration of each visit, and demographics. Massive open online courses (MOOCs) may also provide additional data sets to understand the learning process. For instance, Leony et al. (2015) show how to infer the learners' emotions (i.e., boredom, confusion, frustration, and happiness) by analyzing their actions on the Khan Academy Platform. Moreover, a variety of physiological sensors have been used to increase the quality and depth of analysis (Kaklauskas et al. 2015), such as wearable technologies (Schaefer et al. 2016).

Social computing systems refer to the interplay between people's social behaviors and their interactions with computing technologies (Cheng et al. 2015; Lee and Chen 2013). These systems can extract various kinds of behavioral cues and social signals, such as physical appearance, gesture and posture, gaze and face, vocal behavior, and use of space and environment (Zhou et al. 2012). Analyzing this information can enable the visually representation of social features, such as identity, reputation, trust, accountability, presence, social role, expertise, knowledge, and ownership (Zhou et al. 2012).

There are also open datasets that can be used for research on social and emotional analytics, such as PhysioBank, which includes digital recordings of physiological signals and related data for use by the biomedical research community (Goldberger et al. 2000); DEAP, a database for emotion analysis using physiological signals (Koelstra et al. 2012); and DECAF, a multimodal dataset for decoding user physiological responses to affective multimedia content (Abadi et al. 2015). Verbert et al. (2012) further review the availability of such open educational datasets, including dataTEL (http://www.teleurope.eu/pg/pages/view/50630/), DataShop (https://pslcdatashop.web.cmu.edu/) and Mulce (http://mulce.univ-bpclermont.fr:8080/PlateFormeMulce/). As highlighted by Siemens (2013), taking multiple data sources into account provides more information to educators and students than a single data source.

### Preparation of the data

This step concerns manipulating and converting the raw data materials into suitable forms that will meet the specific input requirements for the DS tools. For example, some DS techniques are designed for symbolic and categorical data, while others handle only numeric values. Typical examples of manipulation include converting data to different types and discretizing or summarizing data to derive new attributes. Moreover, numerical values must often be normalized or scaled so that they are comparable. Preparation also involves sampling, running correlation and significance tests, and data cleaning, which includes removing or inferring missing values. Feature selection and data reduction algorithms may further be used with the cleaned data. The end results are then usually converted to a tabular format for the next step.

Cios and Kurgan (2005) demonstrate that the data preparation step is by far the most time-consuming part of the DS process model, but educational DS research rarely examines this. Cristóbal Romero et al. (2014) survey the literature on pre-processing educational data to provide a guide or tutorial for educators and DS practitioners. Their results showed these seven pre-processing tasks: (1) data gathering, bringing together all the available data into a set of instances; (2) data aggregation/integration, grouping together all the data from different sources; (3) data cleaning, detecting erroneous or irrelevant data and discarding it; (4) user and session identification; identifying individual users; (5) attribute/variable selection, choosing a subset of relevant attributes from all the available attributes; (6) data filtering, selecting a subset of representative data to convert large data sets into smaller data sets; and (7) data transformation, deriving new attributes from the already available ones.

### Mining of the data

At this point, various mining techniques are applied to derive knowledge from preprocessed data (see Table 1). This usually involves the calibration of the parameters to the optimal

**Table 1** Common methods for mining data

| Method | Goal/description | Example tasks | Reference |
|---|---|---|---|
| Classification | To define a set of classes, which are usually mutually exclusive. To predict which classes an individual belongs to. | To automatically detect affective states, like confusion, frustration, and boredom. | Ghergulescu and Muntean (2014) |
| Value estimation | To estimate the numerical value of some variables for an individual. | To estimate learning outcomes regarding student affect and behavioral engagement. | Pardos et al. (2014) |
| Clustering | To measure the similarity of individuals described by data. To group similar individuals together by their similarity, but not driven by any specific purpose | To create groups of students according to their personal characteristics. | He (2013) |
| Frequent pattern mining | To find associations among variables based on their appearing together in transactions and to encode rules. | Identifying relationships in learner behavioral patterns and diagnosing student difficulties. | Kinnebrew et al. (2013) |
| Text mining | To extract high-quality information from text. | Recognize the emotion of interactive text. | Tian et al. (2014) |
| Structural analysis | To predict a link that should exist between individuals, and possibly also estimate the strength of the link. | To dynamically recommend the tutorial dialog in a manner that is responsive to the sensed states. | D'mello and Graesser (2013) |
| Behavior profiling | To characterize the typical or most noticeable behavior of a subgroup or an entire population. | To profile anomalous behaviors. | Hoque and Picard (2014) |

values. The output of this step is some model parameters or pattern capturing regularities in the data.

### Evaluation of the discovered knowledge

The evaluation stage serves to help ensure that the discovered knowledge satisfies the original research goals before moving on. Only approved models are retained for the next step, otherwise the entire process is revisited to identify which alternative actions could be taken to improve the results (e.g., adjusting the problem definition or getting different data). The researchers will assess the results rigorously and thus gain confidence as to whether or not they are qualified. Scheffel et al. (2014) conduct brainstorming with experts from the field of learning analytics and gather their ideas about specific quality indicators to evaluate the effects of learning analytics. We summarize the results in Table 2. The criteria provide a way to standardize the evaluation of learning analytics tools.

In addition, the domain experts will help interpret the results and check whether the discovered knowledge is novel, interesting, and influential. To facilitate their understanding, the research team must think about the comprehensibility of the models to domain experts (and not just to the DS researchers).

As suggested by Romero and Ventura (2010), visualizing models in compelling ways can make analytics data straightforward for non-specialists to observe and understand. For example, Leony et al. (2013) propose four categories of visualizations for an intelligent system, including time-based visualizations, context-based visualizations, visualizations of changes in emotion, and visualizations of accumulated information. The main objective of these visualizations is to provide teachers with knowledge about their learner's emotions, learning causes, and the relationships that learning has with emotions. Verbert et al. (2014) also review works on capturing and visualizing traces of learning activities as dashboard applications. They present examples to demonstrate how visualization can not only promote awareness, reflection, and sense-making, but also represent learner's goals and

**Table 2** Quality indicators for learning analytics. Source: summarized from Scheffel et al. (2014)

| Topic | Criterion | Representative statements |
|---|---|---|
| Objectives | Teacher awareness | Teachers change their behavior in some respects.<br>Teachers react in a more personalized way to how their students are dealing with learning material. |
| | Student awareness | Students become more self-regulated in their learning processes.<br>Students are more aware of their learning progress. |
| Learning Support | Learning support | An early detection of students at risk.<br>The ability to explain what could help them to improve further.<br>Students regularly utilize the tools provided. |
| Learning Measures and Output | Learning outcome | If teachers can gain new insights using the given methods.<br>Results are compared with other (traditional) measures. |
| | Learning performance | Change in workplace learning is measurable.<br>The extent to which the achievement of learning objectives can be demonstrated. |
| Data Aspects | Open access | Data are open access.<br>Portability of the collected data. |
| | Privacy | Privacy is ensured.<br>Learners can influence which data are provided. |
| Organizational Aspects | Acceptance & uptake | Administrators invest in scaling successful tools across their programming. |

enable them to track progress toward these. Epp and Bull (2015) explored 21 visual variables (e.g., arrangement, boundary, connectedness, continuity, depth, motion, orientation, position, and shape) that have been employed to communicate a learner's abilities, knowledge, and interests. Manipulating such visual variables should provide a reasonable starting point from which to visualize educational data.

### Use of the discovered knowledge

This final step consists of planning where and how to put the discovered knowledge into real use. A plan can be obtained by simply documenting the action principles being used to impact and improve teaching, learning, administrative adoption, culture, resource allocation and decision making on investment. The discovered knowledge may also be reported in educational systems, where the learner can see the related visualizations. These visualizations can provide learners with information about several factors, including their knowledge, performance, and abilities (Epp and Bull 2015). Moreover, the results from the current context may be extended to other cases to assess their robustness. The discovered knowledge is then finally deployed.

However, according to the findings of Romero and Ventura (2010) survey, only a small minority of studies can apply the discovered knowledge to institutional planning processes. One of the barriers to this is individual and group resistance to innovation and change. Macfadyen and Dawson (2012) thus highlight that the accessibility and presentation of analytics processes and findings are the keys to motivating participants to feel positive about the change. Furthermore, the initial iteration may not be complete or good enough to deploy, and so a second iteration may be necessary to yield an improved solution. Therefore, the diagram shown in Fig. 1 represents this process as a cycle and describes several explicit feedback loops, rather than as a simple, linear process.

## The case of social-emotional learning

In this section, we describe a case study in which we used the data science research methodology. The research was initiated with an instructor who wanted to understand university students' motivation for learning during a semester. We thus started to help this instructor through understanding the problem (Step 1). The instructor explained that university students' motivation for learning varies over a long semester. Monitoring their motivation can help in providing the right motivated strategies at the right time. We thus went on to the next step: understanding the data (Step 2). Although the use of the motivated strategies for learning questionnaire (MSLQ) (Garcia and Pintrich 1996) can gather data about students' motivation, the questionnaire measures were quite long and were not sensitive to change over time. Inspired by the concept of teaching opinion survey implemented at the end of a semester, we decided to collect text data to evaluate university students' motivation to learn. After repeatedly going through Steps 1 and 2, the research problem became "predicting university students' motivation to learn based on teaching opinion mining."

In this experiment, we employed the motivated strategies for learning questionnaire to collect the respondents' motivation states. In addition, an open-ended opinion survey about the challenges they faced on the F2F course and recommendations to the teacher with regard to adjusting instruction was utilized to collect the text data. One hundred and fifty-two university students (62 females, 90 males; mean age ± S.D. = 21.1 ± 7.5 years) completed the survey for this study. They were taking face-to-face computer courses at four universities in southern Taiwan.

In the data preparation step (Step 3), we first calculated the mean score of MSLQ. Those respondents with a score less than the mean were labeled as low motivation (LM) students, while those with more than the mean were labeled as high motivation (HM) students. The sample consisted of 76 LM and 76 HM students (the mean was equal to the median).

We then continued to process the textual data. Because textual data is unstructured, the aim of data preparation is to represent the raw text by numeric values. This process contained two steps: tokenizing and counting. In the tokenizing step, we used the CKIP Chinese word segmentation system (Ma and Chen 2003) to handle the text segmentation. In the counting step, term frequency-inverse document frequency (TF-IDF) was used as an indicator parameter to extract text features. TF-IDF is a measure of how frequent a term is in a document, and how rare a term is in many documents.

In mining the data (Step 4), we applied a support vector machine (SVM) to classify the respondents. The dataset was randomly split into two groups: a training set and a testing set. The training set consisted of 138 instances (90%) and the testing set of 14 instances (10%). We constructed a model based on the training set and made predictions on the testing set to evaluate the prediction performance. In the evaluation of the model (Step 5), the rate of correct predictions over all instances was measured to represent the accuracy of the prediction model. Through removing the 1074 stop words and substituting the 39 words having similar meanings, the results revealed that the accuracy of the prediction model could be up to 85.7%. We used a free data analysis software, RapidMiner, to perform the analysis (See Fig. 2). Therefore, in the final step the instructor could predict students' motivation to learn during the whole semester using computer-mediated communication, such as instant messaging (Step 6).

We further iterated the process by redefining the research problem as "finding groups of respondents using similar terms to describe an opinion." In mining the data, the K-Means clustering method was used to partition the respondents into two clusters. The cluster model revealed that Cluster 1 had 89 respondents, and Cluster 2 had 63. ANOVA was performed to determine how the score of MSLQ was influenced by participant's

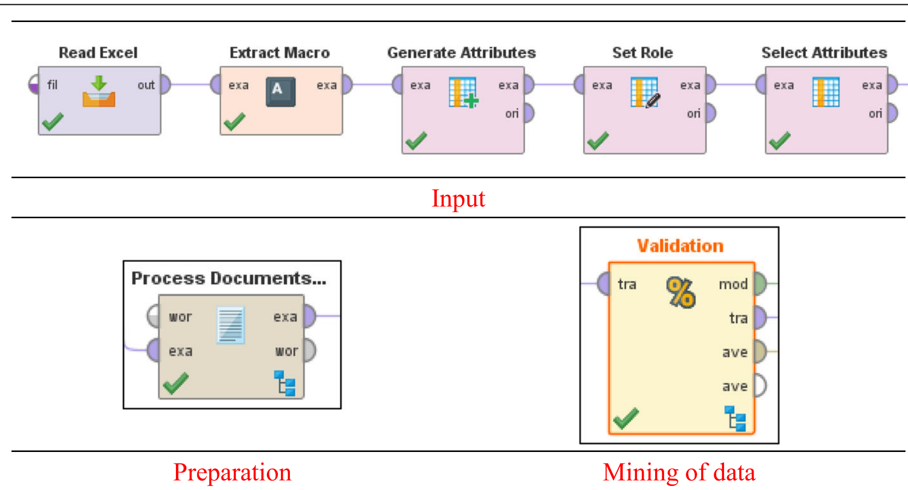


**Fig. 2** The analysis process in RapidMiner

**Table 3** Mean scores, standard deviations, and analysis of variance (ANOVA) results for both clusters

| clusters | N | MSLQ | | | ANOVA | |
|---|---|---|---|---|---|---|
| | | Mean | SD | 95% CI to mean | F(1, 150) | Cohen's d |
| cluster 1 | 89 | 106.43 | 12.55 | 103.78 to 109.07 | 14.33*** | 0.29 |
| cluster 2 | 63 | 114.78 | 14.52 | 111.12 to 118.43 | | |

***$p < .001$

clusters (see Table 3). Significant effects across different work methods were found for the two clusters, F (1, 150) = 14.33, $p = .000$. Table 3 indicates that the Cluster 2 had a higher mean score of MSLQ than Cluster 1. The cluster model also found that the top three important terms for were "考試(exam)", "報告(presentation)", and "作業(homework)" for Cluster 1 and "老師(instructor)", "同學(peer)", and "自己(oneself)" for Cluster 2. In other words, the terms used in Cluster 1 concerned more about the value component of MSLQ. However, the terms used in Cluster 2 concerned more about the expectancy component of MSLQ. Therefore, the instructor could use these terms to roughly provide interventions to improve students' motivation for learning.

## Conclusion

The broad availability of data has led to the development of data science. This paper's research goals are to stimulate further research and practice in the use of data science for education. It also presents a DS research methodology that is applicable to achieve these goals. A well-defined DS research model can help ensure that quality of results, contribute to better understanding the techniques behind the model, and lead to faster, more reliable, and more manageable knowledge discovery. Through an examination of large data sets, a DS methodology can help us to acquire more knowledge about how people learn (Koedinger et al. 2015). This is important, as it contributes to the development of better intervention support for more effective learning.

This paper also describes the emerging field of social-emotional learning and its challenges. It has been proposed that the social-emotional competencies that occur between people will become very important to education in the future. Although research suggests that social-emotional qualities have a positive influence on academic achievement, most related studies examine these qualities in relation to outcome measurement and prediction, and more work is needed to develop interventions based on this research (Levin 2013). Therefore, this paper presents a case study of social-emotional learning in which we used the data science research methodology.

Several large problems remain to be addressed by researchers in this field. Before incorporating the approaches recommended in this work in large-scale education settings, we should select a few social-emotional skill areas and measures. This investment in data acquisition and knowledge discovery by DS will enable a deeper understanding of school effects and school policy in this context, and would avoid pulling reform efforts in unproductive or detrimental directions (Whitehurst 2016). Moreover, explicit privacy regulations, such as anonymity in data collection and consent from the parents in a K-12 setting, also need to be addressed. Slade and Prinsloo (2013) recommend collaborating with students on voluntarily providing data and allowing them to access DS outcomes to aid in their learning and development. We hope the issues we have highlighted in this paper help stimulate further research and practice in education.

**Acknowledgements**
This research is partially supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. MOST 105-2511-S-006 -015 -MY2.

**Authors' contributions**
Both authors read and approved the final manuscript.

**Competing interests**
The authors declare that they have no competing interests.

Received: 21 September 2016 Accepted: 12 December 2016
Published online: 13 January 2017

**References**
M.K. Abadi, R. Subramanian, S.M. Kia, P. Avesani, I. Patras, N. Sebe, DECAF: MEG-based multimodal database for decoding affective physiological responses. IEEE Trans. Affect. Comput. **6**(3), 209–222 (2015). doi:10.1109/taffc.2015.2392932
B. Baumer, A data science course for undergraduates: Thinking with data. Am. Stat. **69**(4), 334–342 (2015). doi:10.1080/00031305.2015.1081105
M.A. Brackett, S.E. Rivers, M.R. Reyes, P. Salovey, Enhancing academic performance and social and emotional competence with the RULER feeling words curriculum. Learn. Individ. Differ. **22**(2), 218–224 (2012). doi:10.1016/j.lindif.2010.10.002
Q. Cheng, X. Lu, Z. Liu, J.C. Huang, Mining research trends with anomaly detection models: The case of social computing research. Scientometrics **103**(2), 453–469 (2015). doi:10.1007/s11192-015-1559-9
Child Trends, *Measuring Elementary School Students' Social and Emotional Skills: Providing Educators with Tools to Measure and Monitor Social and Emotional Skills that Lead to Academic Success*, 2014. Retrieved from http://www.childtrends.org/wp-content/uploads/2014/08/2014-37CombinedMeasuresApproachandTablepdf1.pdf
K.J. Cios, L.A. Kurgan, Trends in data mining and knowledge discovery, in *Advanced techniques in knowledge discovery and data mining*, ed. by N.R. Pal, L. Jain (Springer London, London, 2005), pp. 1–26
K.J. Cios, R.W. Swiniarski, W. Pedrycz, L.A. Kurgan, *The knowledge discovery process data mining: A knowledge discovery approach* (Springer US, Boston, 2007), pp. 9–24
G. Cobb, Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. Am. Stat. **69**(4), 266–282 (2015). doi:10.1080/00031305.2015.1093029
S. D'mello, A. Graesser, AutoTutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. ACM Trans. Interac. Intell. Sys. **2**(4), 1–39 (2013). doi:10.1145/2395123.2395128
V. Dhar, Data science and prediction. Commun. ACM **56**(12), 64–73 (2013). doi:10.1145/2500499
A.L. Duckworth, D.S. Yeager, Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. Educ. Res. **44**(4), 237–251 (2015). doi:10.3102/0013189x15584327
C.D. Epp, S. Bull, Uncertainty representation in visualizations of learning analytics for learners: Current approaches and opportunities. IEEE Trans. Learn. Technol. **8**(3), 242–260 (2015). doi:10.1109/tlt.2015.2411604
A. Fernández, D. Peralta, J.M. Benítez, F. Herrera, E-learning and educational data mining in cloud computing: An overview. Int. J. Learn. Technol. **9**(1), 25–52 (2014). doi:10.1504/IJLT.2014.062447
T. Garcia, P. Pintrich, Assessing Students' motivation and learning strategies in the classroom context: the motivated strategies for learning questionnaire, in *Alternatives in assessment of achievements, learning processes and prior knowledge*, ed. by M. Birenbaum, F.R.C. Dochy, vol 42 (Springer, Netherlands, 1996), pp. 319–339
P.G. Ghazarian, S. Kwon, The future of American education: Trends, strategies, & realities. Philos. Educ. **56**, 147–177 (2015)
I. Ghergulescu, C.H. Muntean, A novel sensor-based methodology for learner's motivation analysis in game-based learning. Interact. Comput. **26**(4), 305–320 (2014). doi:10.1093/iwc/iwu013
A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, H.E. Stanley, PhysioBank, PhysioToolkit, and PhysioNet - components of a new research resource for complex physiologic signals. Circulation **101**(23), E215–E220 (2000)
H. Guruler, A. Istanbullu, Modeling student performance in higher education using data mining, in *Educational data mining: applications and trends*, ed. by A. Peña-Ayala (Springer International Publishing, Cham, 2014), pp. 105–124
J. Hardin, R. Hoerl, N.J. Horton, D. Nolan, B. Baumer, O. Hall-Holt, P. Murrell, R. Peng, P. Roback, D.T. Lang, M.D. Ward, Data science in statistics curricula: Preparing students to "think with data". Am. Stat. **69**(4), 343–353 (2015). doi:10.1080/00031305.2015.1077729
W. He, Examining students' online interaction in a live video streaming environment using data mining and text mining. Comput. Hum. Behav. **29**(1), 90–102 (2013). doi:10.1016/j.chb.2012.07.020
JJ Heckman, T Kautz, Fostering and measuring skills: interventions that improve character and cognition. National Bureau of Economic Research Working Paper Series, 19656 (2013). doi:10.3386/w19656
M. Hoque, R.W. Picard, Rich nonverbal sensing technology for automated social skills training. Computer **47**(4), 28–35 (2014)
Y-M Huang, M-C Liu, C-H Lai, C-J Liu, Using humorous images to lighten the learning experience through questioning in class. Br. J. Educ. Technol. (In Press). doi:10.1111/bjet.12459
A. Kaklauskas, A. Kuzminske, E. K. Zavadskas, A. Daniunas, G. Kaklauskas, M, Seniut …R. Cerkauskiene (2015). Affective tutoring system for built environment management. Computers & Education, **82**, 202–216. doi:10.1016/j.compedu.2014.11.016
A. Kamenetz, *Nonacademic skills are key to success. But what should we call them?* 2015. Retrieved from National Public Radio website: http://www.npr.org/sections/ed/2015/05/28/404684712/non-academic-skills-are-key-to-success-but-what-should-we-call-them
J.S. Kinnebrew, K.M. Loretz, G. Biswas, A contextualized, differential sequence mining method to derive students' learning behavior patterns. J. Educ. Data Min. **5**(1), 190 (2013)

K.R. Koedinger, S. D'Mello, E.A. McLaughlin, Z.A. Pardos, C.P. Rose, Data mining and education. Wiley Interdiscip. Rev. Cogn. Sci. **6**(4), 333–353 (2015). doi:10.1002/wcs.1350

S. Koelstra, C. Muehl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, DEAP: A database for emotion analysis using physiological signals. IEEE Trans. Affec. Comput. **3**(1), 18–31 (2012). doi:10.1109/t-affc.2011.15

C.-H. Lai, M.-C. Liu, C.-J. Liu, Y.-M. Huang, Using positive visual stimuli to lighten the online synchronous learning experience through in-class questioning. Int. Rev. Res. Open Distance Learn. **17**(1), 23–41 (2016). doi:10.19173/irrodl.v17i1.2114

M.R. Lee, T.T. Chen, Understanding social computing research. It Professional **15**(6), 56–62 (2013)

D. Leony, P.J. Munoz-Merino, A. Pardo, C.D. Kloos, Provision of awareness of learners' emotions through visualizations in a computer interaction-based environment. Expert. Sys. App. **40**(13), 5093–5100 (2013). doi:10.1016/j.eswa.2013.03.030

D. Leony, P.J. Munoz-Merino, J.A. Ruiperez-Valiente, A. Pardo, C.D. Kloos, Detection and evaluation of emotions in massive open online courses. J. Universal. Comput. Sci. **21**(5), 638–655 (2015)

H.M. Levin, The utility and need for incorporating noncognitive skills into large-scale educational assessments, in *The role of international large-scale assessments: perspectives from technology, economy, and educational research*, ed. by M. von Davier, E. Gonzalez, I. Kirsch, K. Yamamoto (Springer Netherlands, Dordrecht, 2013), pp. 67–86

L.C. Linan, A.A.J. Perez, Educational data mining and learning analytics: Differences, similarities, and time evolution. Rusc-Univ. Knowl. Soc. J. **12**(3), 98–112 (2015). doi:10.7238/rusc.v12i3.2515

E. Lindqvist, R. Vestman, The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. Am. Econ. J. Appl. Econ. **3**(1), 101–128 (2011). doi:10.1257/app.3.1.101

A.A. Lipnevich, R.D. Roberts, Noncognitive skills in education: Emerging research and applications in a variety of international contexts. Learn. Individ. Differ. **22**(2), 173–177 (2012). doi:10.1016/j.lindif.2011.11.016

C.-J. Liu, C.-F. Huang, M.-C. Liu, Y.-C. Chien, C.-H. Lai, Y.-M. Huang, Does gender influence emotions resulting from positive applause feedback in self-assessment testing? Evidence from neuroscience. Educ. Technol. Soc. **18**(1), 337–350 (2015)

W.-Y. Ma, K.-J. Chen, *A bottom-up merging algorithm for Chinese unknown word extraction*, 2003. Paper presented at the second SIGHAN workshop on Chinese language processing, Sapporo, Japan

L.P. Macfadyen, S. Dawson, Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. Educ. Technol. Soc. **15**(3), 149–163 (2012)

I. Mendez, The effect of the intergenerational transmission of noncognitive skills on student performance. Econ. Educ. Rev. **46**, 78–97 (2015). doi:10.1016/j.econedurev.2015.03.001

T.E. Moffitt, L. Arseneault, D. Belsky, N. Dickson, R.J. Hancox, H. Harrington, R. Houts, R. Poulton, B.W. Roberts, S. Ross, M. R. Sears, W.M. Thomson, A. Caspi, A gradient of childhood self-control predicts health, wealth, and public safety. Proc. Natl. Acad. Sci. U. S. A. **108**(7), 2693–2698 (2011). doi:10.1073/pnas.1010076108

KA Moore, LH Lippman, R Ryberg, Improving outcome measures other than achievement. AERA Open, 1(2) (2015). doi:10.1177/2332858415579676

Z. Papamitsiou, A.A. Economides, Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. Educ. Technol. Soc. **17**(4), 49–64 (2014)

ZA Pardos, RSJD Baker, MS Pedro, SM Gowda, SM Gowda, Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. J. Lear. Anal. (2014)

F. Provost, T. Fawcett, Data science and its relationship to big data and data-driven decision making. Big Data **1**(1), 51–59 (2013). doi:10.1089/big.2013.1508

F Provost, T Fawcett, Data Science for Business: What you need to know about data mining and data-analytic thinking (Sebastopol, CA: O'Reilly Media, Inc, 2013)

X.L. Ren, A research on future education development strategies in China. Philosophy of Education **56**, 69–118 (2015)

C. Romero, S. Ventura, Educational data mining: a review of the state of the art. IEEE Transact. Sys. Man. Cybern. Part C-Appl. Rev. **40**(6), 601–618 (2010). doi:10.1109/tsmcc.2010.2053532

C. Romero, J.R. Romero, S. Ventura, A survey on pre-processing educational data, in *Educational data mining: applications and trends*, ed. by A. Peña-Ayala (Springer International Publishing, Cham, 2014), pp. 29–64

S.E. Schaefer, C.C. Ching, H. Breen, J.B. German, Wearing, thinking, and moving: Testing the feasibility of fitness tracking with urban youth. Am. J. Health Educ. **47**(1), 8–16 (2016). doi:10.1080/19325037.2015.1111174

M. Scheffel, H. Drachsler, S. Stoyanov, M. Specht, Quality indicators for learning analytics. Educ. Technol. Soc. **17**(4), 117–132 (2014)

S. Sellar, Data infrastructure: A review of expanding accountability systems and large-scale assessments in education. Discourse **36**(5), 765–777 (2015). doi:10.1080/01596306.2014.931117

C. Shearer, The CRISP-DM model: The new blueprint for data mining. J. Data Warehousing **5**(4), 13–22 (2000)

G. Siemens, Learning analytics: The emergence of a discipline. Am. Behav. Sci. **57**(10), 1380–1400 (2013). doi:10.1177/0002764213498851

S. Slade, P. Prinsloo, Learning analytics: Ethical issues and dilemmas. Am. Behav. Sci. **57**(10), 1510–1529 (2013). doi:10.1177/0002764213479366

S.D. Sparks, *Scholars: better gauges needed for 'mindset,' 'grit' retrieved from education week website*, 2016. http://www.edweek.org/ew/articles/2016/04/20/scholars-better-gauges-needed-for-mindset-grit.html

F. Tian, P.D. Gao, L.Z. Li, W.Z. Zhang, H.J. Liang, Y.A. Qian, R.M. Zhao, Recognizing and regulating e-learners' emotions based on interactive Chinese texts in e-learning systems. Knowl.-Based Syst. **55**, 148–164 (2014). doi:10.1016/j.knosys.2013.10.019

K. Verbert, N. Manouselis, H. Drachsler, E. Duval, Dataset-driven research to support learning and knowledge analytics. Educ. Technol. Soc. **15**(3), 133–148 (2012)

K. Verbert, S. Govaerts, E. Duval, J.L. Santos, F. Van Assche, G. Parra, J. Klerkx, Learning dashboards: An overview and future research opportunities. Pers. Ubiquit. Comput. **18**(6), 1499–1514 (2014). doi:10.1007/s00779-013-0751-2

G.J. Whitehurst, *Hard thinking on soft skills*, 2016. Retrieved from Brookings Institution, http://www.brookings.edu/research/reports/2016/03/24-hard-thinking-soft-skills-whitehurst

World Economic Forum, *The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution*, 2016. Retrieved from World Economic Forum, http://www3.weforum.org/docs/Media/WEF_Future_of_Jobs_embargoed.pdf

J.H. Zhou, J.Z. Sun, K. Athukorala, D. Wijekoon, M. Ylianttila, Pervasive social computing: augmenting five facets of human intelligence. J. Ambient. Intell. Humaniz. Comput. **3**(2), 153–166 (2012). doi:10.1007/s12652-011-0081-z