A Delphi Study of Competencies and Skills for Data Analytics (WIP)

Orli Weiser
orliwe@openu.ac.il
Ben-Gurion University of the Negev

Yoram M. Kalman
yoramka@openu.ac.il
The Open University of Israel

Carmel Kent
kent.carmel@gmail.com
Arden University, UK

Gilad Ravid
rgilad@bgu.ac.il
Ben-Gurion University of the Negev

Abstract

The ability to collect digital records, analyze them and effectively communicate the findings is emerging as a key competency set for graduates of management programs and of other social science programs. This work-in-progress extended abstract describes a Delphi study of the competencies needed for effective data analytics. Fifty international experts will be invited to participate in the study and share their opinion regarding cognitive, intrapersonal and interpersonal competencies. The output collected will be analyzed using "Q methodology". The results of this study will serve us in developing a data-analytics student lab that will support the development of these competencies and improve student readiness for the future labor market.

Keywords: big data, data analytics, competencies, Delphi technique, Q methodology, CRISP-DM model, virtual lab

I. INTRODUCTION

Recent developments in digital technologies open up exciting opportunities for social science research and for organizational decision making. Labels such as "big data" describe the potential of analyzing digital records created on and collected by information systems. The ability to collect digital records, analyze them and effectively communicate the findings is emerging as a key competency set for researchers in the social sciences, and for practitioners in organizations of every size.

Thus, teaching data literacy skills is an important goal for academic institutions.

In order to deal with these challenges, we are developing a data-analytics student lab. This project will provide undergraduates and graduates students with a virtual "lab bench" where they can access datasets, analysis and visualization tools, as well as structured methodologies on how to carry out lab experiments and observational studies. This lab will incorporate elements from virtual labs used in science education (Heradio et. al., 2016). The lab will enable faculty to develop data analytics lab exercises, allow students to carry out these experiments/analyses, allow researchers to track and monitor student activities and apply learning analytics to assess each student's performance and to provide personalized feedback.

In the course of developing this student lab, we were confronted with the question of what are the learning outcomes this lab is supposed to support? A review of the literature revealed the lack of a contemporary conceptualization of the skills and capabilities required for data analytics in the context of computational social science (Lazer et al., 2009). Hence we set out to develop a description of this competency set by conducting a Delphi study focusing on the competencies needed for data analytics.

One key resource for this conceptualization are "21st century skills" frameworks (e.g. Eshet, 2004; Pellegrino & Hilton, 2012). These frameworks include cognitive, intrapersonal and interpersonal competencies required for being productive contributing citizens and professionals in the current and future knowledge society. Additional resources are the AIS "IS 2010 Curriculum Guidelines" (Topi et al., 2010) and the "Innovating Pedagogy" report which explores new forms of teaching, learning and assessment for an interactive world (Ferguson et al., 2017).

II. METHOD

The Delphi technique is a structured process that uses a series of questionnaire-based rounds to gather information from a chosen panel of respondents within their domain of expertise, until group consensus is determined to be achieved (Dalkey & Helmer, 1963). In the traditional Delphi study, the first questionnaire is based on an extensive review of the literature. During the years, additional applications of the Delphi technique were developed. The most popular of them is the Modified Delphi technique, which is similar to the traditional Delphi technique in terms of procedure and intent, but it can involve interviews and group interaction and ignore the anonymity rule kept in the traditional Delphi technique (Custer, Scarcella & Stewart, 1999). The modified technique is applicable for goal settings, for policy investigation and for predicting future events (Hsu & Sandford, 2007).

The current study will be based on the Modified Delphi technique. The first round will include a questionnaire developed by the research team, based on a literature review and on exploratory interviews of a small number of experts. Following questionnaire development, 50 experts will be invited to participate in the study. The anonymity of the experts will be kept, and their identity will be known to the research team only. The experts will be chosen according to a list of requirements described in Appendix A. The Experts will be recruited from industry and academia, and their expertise will represent all six phases of data analytics according to the CRISP-DM model [a "Cross-Industry Process for Data Mining" model that breaks the process of data mining into six major phases: Business understanding, Data understanding, Data preparation, Modeling, Evaluation and Deployment (Chapman et al., 2000)].

The questionnaire that will be sent to the experts in the first round will detail cognitive, intrapersonal and interpersonal competencies from the Pellegrino & Hilton

model (Pellegrino & Hilton, 2012), as well as additional competencies identified in the literature described in the Introduction section. The respondents will be asked to rate the importance of each competency both for novice data analysts and for expert data analysts in the foreseeable future (5-10 years), by dragging and dropping each competency into an appropriate "box" ("Essential competency", "Important competency", "Less important competency", "Not an important competency", "I do not know", "I do not understand this item"). The items will be accompanied by open text questions, in which the experts will be able to elaborate on their responses, as well as add their own opinion and suggestions.

The research team will analyze the replies to each round's questionnaire using "Q methodology". The "Q methodology" is a research methodology within social science in which the individuals perform the measuring rather than being measured (McKeown & Thomas, 2013). It is based on qualitative but statistical approach which enables discovering discourses concerning how individuals understand the social and environmental worlds in which they live (Barry & Proops, 1999).The respondents' ranking of the items (Q sort) will reflect the individual's response to this item and will be statistically analyzed using factor analysis methods.

The qualitative data will be analyzed too and an updated questionnaire will be sent to the participating experts asking for their feedback to the other experts' ratings and suggestions, until a consensus will be reached. The final report will include the ranked items, and if necessary, a list of items for which a consensus could not reached.

III. SUMMARY

The Delphi study will result in ranked lists describing the consensus among data analytics experts regarding the competency set for novice data analysts and expert data analysts. These will inform our development of the data-analytics student lab and of lab assignments for the students.

REFERENCES

Barry, J., & Proops, J. (1999). Seeking sustainability discourses with Q methodology. *Ecological economics*, 28(3), 337-345.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*. Technical report, CRISP-DM.

Custer, R. L., Scarcella, J. A., & Stewart, B. R. (1999). The modified Delphi technique-A rotational modification. Journal of Career and Technical Education, 15(2).

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management science*, 9(3), 458-467.

Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia and Hypermedia*, *13*(1), 93-107.

Ferguson, R., Barzilai, S., Ben-Zvi, D., Chinn, C.A., Herodotou, C., Hod, Y., Kali, Y., Kukulska-Hulme, A., Kupermintz, H., McAndrews, P., Rienties, B., Sagy, O., Scanlon, E., Sharples, M., Weller, M. & Whitelock, D. (2017). *Innovating Pedagogy 2017*. OpenUniv. Retrieved January 14, 2018 from https://www.learntechlib.org/p/182004/.

Heradio, R., de la Torre, L., Galan, D., Cabrerizo, F. J., Herrera-Viedma, E., & Dormido, S. (2016). Virtual and remote labs in education: A bibliometric analysis. *Computers & Education*, 98, 14-38.]

Hsu, C. C., & Sandford, B. A. (2007). The Delphi technique: making sense of consensus. *Practical assessment, research & evaluation*, *12*(10), 1-8.

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D. & Jebara, T. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY),* 323(5915), 721.

McKeown, B., & Thomas, D. B. (2013). *Quantitative Applications in the Social Sciences Series: Vol. 66.Q methodology (2nd ed.).* Thousand Oaks, CA: Sage publications.

Pellegrino J.W., & Hilton, M. L. (2012). Education for life and work: Developing transferable knowledge and skills in the 21st century. *Committee on Defining Deeper Learning and 21st Century Skills. Division on Behavioral and Social Sciences and Education. National Research Council. Washington, DC: National Research Council*.

Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K., Nunamaker Jr, J. F., Sipior, J. C., & de Vreede, G. J. (2010). IS 2010: Curriculum guidelines for undergraduate degree programs in information systems. *Communications of the Association for Information Systems*, 26(1), 18.

APPENDIX A - EXPERT CHARACTERISTICS

Table 1. Expert characteristics

Criteria	Requirements
Number of experts	Minimum of 50
Place of work and research	Minimum of 50% International (worked/researched in more than one country)
Field of work	A diverse selection of industry, governmental and non-governmental organizations, and academia. Minimum of 15% for each sector
Gender	At least 30% for each gender
Education	Relevant for the domain of expertise. No formal degree required
Professional experience	Minimum of 3 years experience in data analytics. The average experience in data analytics of all participants will be at least 5 years. At least 20% will have some managerial/decision level experience in the field.
Domain of expertise in data analytics/data science	Relevant to the study. At least one expert from each phase of the CRISP-DM model
Willing to commit to responding to several rounds of online questionnaires	Mandatory