New Proposal to Compare Student Data in Institutional Research

○Yasuo NAKATA
Faculty of Health Sciences,
Organization for the Advancement of Higher Education,
Center for the Promotion of Excellence in Research and Development of Higher Education
Kobe Tokiwa University,
Kobe, Japan
nakata0325@gmail.com

Katsuhiko MURAKAMI
Human Genome Center,
The Institute of Medical Science,
The University of Tokyo,
Tokyo, Japan
murakami.ktk@gmail.com

Yasuhiro KOZAKI
Home Economics Education,
Department of Education,
Osaka Kyoiku University,
Osaka, Japan

The Center for Early Childhood Development, Education, and Policy Research,
The University of Tokyo,
Tokyo, Japan
kozaki@cc.osaka-kyoiku.ac.jp

Takafumi Kirimura
Faculty of Education,
Hirosaki University
kirimura@hirosaki-u.ac.jp

Aoi SUGIURA
Kobe City Nishi-Kobe Medical Center,
Kobe, Japan
aoi.kishida@gmail.com

Kenya BANNAKA
Department of Oral Health
Center for the Promotion of Excellence in Research and Development of Higher Education
Kobe Tokiwa College,
Kobe, Japan
k-bannaka@kobe-tokiwa.ac.jp

Kunihiko TAKAMATSU*
Faculty of Education,
Organization for the Advancement of Higher Education,
Center for the Promotion of Excellence in Research and Development of Higher Education,
Life Science Center
Kobe Tokiwa University,
Kobe, Japan
ktakamatu@gmail.com
*Corresponding Author

***Abstract*—This article proposes new criteria for using student data in universities. First criteria are called primary data and secondary criteria are called secondary data.**

We define primary data as those that are not linear combination data, and secondary data as a linear combination of primary data. For example, at the macro-level, primary data are correct and incorrect answers to a question in an examination or students' attendance and absence from a lecture. At the macro-level, secondary data are the total points in an examination or students' total attendance in and absence from a lecture. At the meso-level, secondary data are student records of lectures as well as grade point average, or rank, in the annual record of the university.

Primary data are mainly constructed by faculty while secondary data are constructed by administrative staff. To compare primary and secondary data, collaboration between faculty and administrative staff is important.

Keywords—primary data, secondary data, institutional research, faculty and staff collaboration, linear function combination

I. Introduction

Recently, we proposed a new field called Eduinformatics [1]. It is a coined word, combining "education" and "informatics," which is similar to Bioinformatics, a combination of "biology" and "informatics." In our previous article, we explained that it is important to create new methods in informatics to analyze educational evaluation.

To deal with student data, universities have a department called institutional research (IR). The Ministry of Education, Culture, Sports, Science and Technology (MEXT) published two famous reports on IR [2], [3]. The number of IR offices in Japanese universities increased after these reports were published. Since 2016, MEXT has been strongly requiring Japanese universities to construct IR departments.

In Japan, the International Conference on Data Science and Institutional Research (DSIR) is the only international meeting that deals with IR. Since 2016, we have published some articles in DSIR [4]–[8] that address the topics of collaboration research between faculty and staff and the application of bioinformatics in education.

In the present article, we proposed new criteria for using student data in universities or IR.

II. New criteria

The United Nations Educational, Scientific and Cultural Organization (UNESCO) proposed "three layers of learning analytics" in a report on learning analytics in 2012 [9] (Figure 1). The following are the definitions of the three layers of analytics from [9].

"Macro-level analysis seek to enable cross- institutional analytics, for instance, through 'maturity' surveys of current institutional practices[10] or improving state-wide data access to standardized assessment data over students' lifetimes[11]. Macro-analytics will become increasingly real-time, incorporating more data from the finer-granularity meso/micro levels, and could conceivably benefit from benchmarking and data integration methodologies developed in non-educational sectors (although see below for concerns about the dangers of decontextualized data and the educational paradigms they implicitly perpetuate).

Meso-level analytics operate at institutional level. To the extent that educational institutions share common business processes to sectors already benefiting from Business Intelligence, BI, they can be seen as a new BI market sector, who can usefully appropriate tools to integrate data silos in enterprise warehouses, optimize workflows, generate dashboards, mine unstructured data, better predict 'customer churn' and future markets, and so forth. It is the BI imperative to optimize business processes that partly motivates efforts to

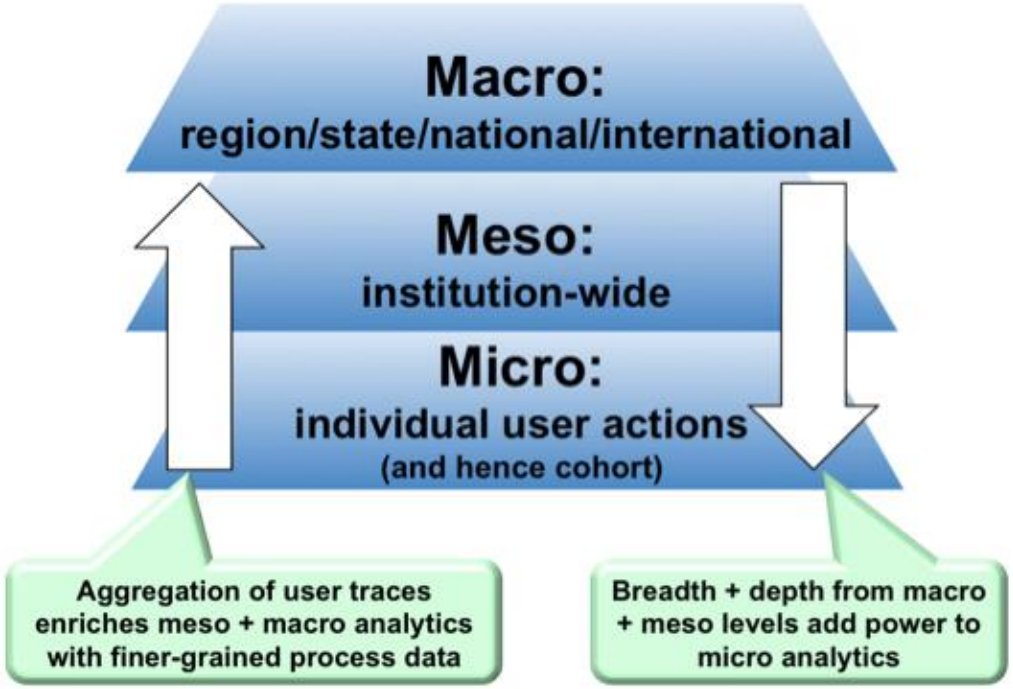

Figure 1 Layers of Learning Analytics from UNESCO IITE, Learning Analytics, 2012

build institutional-level "academic analytics"[12], and we see communities of practice specifically for BI within educational organizations, which have their own cultures and legacy technologies.

Micro-level analytics support the tracking and interpretation of process-level data for individual learners (and by extension, groups). This data is of primary interest to learners themselves, and those responsible for their success, since it can provide the finest level of detail, ideally as rapidly as possible. This data is correspondingly the most personal, since (depending on platforms) it can disclose online activity click-by-click, physical activity such as geolocation, library loans, purchases, and interpersonal data such as social networks. Researchers are adapting techniques from fields including serious gaming, automated marking, educational data mining, computer-supported collaborative learning, recommender systems, intelligent tutoring systems/adaptive hypermedia, information visualization, computational linguistics and argumentation, and social network analysis."

In micro-level analytics, we deal with two types of data, primary data and secondary data.

We define a linear function combination as follows.

For function $f_1, f_2, \ldots, f_n$ and scalar $a_1, a_2, \ldots, a_n$, we define the linear combination of $f_1, f_2, \ldots, f_n$, as

$$a_1 f_1 + a_2 f_2 + \ldots + a_n f_n.$$

Therefore, a linear combination is the summation of the function f times weight $a_1, a_2, \ldots, a_n$.

Further, we define a linear map or function as follows.

For function or map x and y, the linear map or function satisfy the following two properties.

Additivity: $f(x + y) = f(x) + f(y)$.

Homogeneity of degree 1: $f(ax) = a f(x)$ for all a.

Obviously, when a map or function is linear, it is a linear combination.

Now, we define primary data as those data that are not linear combination data, while secondary data are a linear combination of primary data.

For example, at the macro-level, primary data are correct and incorrect answers to a question in an examination or students' attendance in and absence from a lecture. On the other hand, at the macro-level, secondary data are the total points in an examination or students' total attendance in and absence from a lecture.

In addition, at the meso-level, for example, secondary data are student records of lectures as well as grade point average (GPA) or rank in the annual record of university. GPA and rank in the annual record of the university, meaning a linear function combination, are calculated using student record of lectures, which are secondary data. Remarkably, therefore, secondary data are calculated using not only primary data but also secondary data.

III. Summary and Future Works

Researchers or IR staff only compare primary data at the micro-level and secondary data at the meso-level in universities. How can we compare between primary and secondary data in universities? Recent research has provided a good suggestion.

First, we must consider who constructs primary and secondary data in universities. Usually, primary data at the micro-level are only obtained and stored by university faculty. On the other hand, the entire secondary data set at the meso-level is only stored by IR staff in universities. This means that faculty cannot access the entire secondary data set, and staff cannot access primary data.

In recent years, in Japanese universities, faculty evaluate students based on rubric. Faculty evaluate not only examinations but also the attitude of students who participate in lectures. This means that student record of lectures does not constitute primary data at the micro-level. Student record data are secondary data and are calculated as a linear function combination of both examination points and attitude points based on rubric. In this case, a specific rubric point of a student constitutes primary data at the micro-level.

Further, we should consider examination points. Usually, an examination is a combination of questions. Most faculty do not record students' correct or incorrect answers to each question. In this case, students' correct or incorrect answers for each question are primary data. Examination points are secondary data at the micro-level. As can be easily imagined, most faculty do not record such primary data. To record such primary data, faculty must use online examinations or computer-scored answer sheets.

Moreover, to compare primary and secondary data, faculty give these data to the IR department. Additionally, at the same time, the IR department prepares and stores primary data constructed by faculty. The most important thing, as per previous research, in comparing primary and secondary data is collaboration between faculty and staff in universities, similar to our research [4]–[8].

Second, the use of Information and Communication Technology in education gave rise to the possibility of

comparing primary data at the micro-level and primary or secondary data at the meso-level in universities. In fact, a Learning Management System (LMS), for instance, massive open online course (MOOC), constructs many primary data at the micro-level. For example, login and time of learning data are primary data at the micro-level. Recently, Kondo et al. compared LMS log data, that is, primary data, and other secondary data [13].

We have already obtained correct and incorrect data for each question on mathematics ability of first-year students at Japanese universities since 2012. In the future, we will compare these primary and secondary data, for example, GPA or rank in the annual record of the university. In these cases, there are too many elements of the question to compare between primary and secondary data. To reduce the number of elements of primary data, we may have to utilize Item Response Theory.

REFERENCES

[1] K. Takamatsu, K. Murakami, T. Kirimura, K. Bannaka, I. Noda, L. R.-J. Wei, K. Mitsunari, M. Seki, E. Matsumoto, M. Bohgaki, A. Imanishi, M. Omori, R. Adachi, M. Yamasaki, H. Sakamoto, K. Takao, J. Asahi, T. Nakamura, and Y. Nakata, "'Eduinformatics': A new education field promotion," *Bull. kobe Tokiwa Univ.*, vol. 11, pp. 27–44, 2018.

[2] MEXT, "Towards the qualitative transformation of university education in order to build a new future," *The Central Council for Education*, 2012. [Online]. Available: http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1325047.htm. [Accessed: 15-Mar-2018].

[3] MEXT, "Construction of education for university," *The Central Council for Education*, 2008. [Online]. Available: http://www.mext.go.jp/b_menu/shingi/chukyo/chukyo4/houkoku/080410.htm. [Accessed: 15-Mar-2018].

[4] K. Takamatsu, T. Kirimura, K. Bannaka, I. Noda, M. Omori, R. Adachi, K. Mitsunari, T. Nakamura, and Y. Nakata, "SWOT analysis and complex network analysis to enhance governance in universities by collaboration between faculty and staff," in *Advanced Applied Informatics (IIAI-AAI), 2016 5th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2016, pp. 1188–1189.

[5] K. Takamatsu, K. Murakami, T. Kirimura, K. Bannaka, I. Noda, M. Yamasaki, L. Rahpael-Joe, l Wei, K. Mitsunari, T. Nakamura, and Y. Nakata, "A new way of visualizing curricula using competencies: Cosine similarity, multidimensional scaling methods, and scatter plotting," in *Advanced Applied Informatics (IIAI-AAI), 2017 6th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2017, pp. 192–197.

[6] M. Seki, Y. Nakata, Y. Kozaki, K. Murakami, E. Matsumoto, K. Bannaka, I. Noda, K. Mitsunari, and K. Takamatsu, "Generic Support System Prototype in Eduinformatics for Students Selecting a Seminar," in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 470–473.

[7] K. Murakami, K. Takamatsu, Y. Kozaki, A. Kishida, K. Bannaka, I. Noda, J. Asahi, K. Takao, K. Mitsunari, T. Nakamura, and Y. Nakata, "Predicting the Probability of Student Dropout through EMIR Using Data from Current and Graduate Students," in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 478–481.

[8] K. Takamatsu, Y. Kozaki, K. Murakami, K. Bannaka, I. Noda, W. Lim, Raphael-Joel, K. Mitsunari, T. Nakamura, and Y. Nakata, "A New Way of Visualizing Curricula Using Competencies : Cosine Similarity and t-SNE," in *Advanced Applied Informatics (IIAI-AAI), 2018 7th International Institute of Applied Informatics (IIAI) International Congress on. Institute of Electrical and Electronics Engineers (IEEE)*, 2018, pp. 390–395.

[9] UNESCO, "Learning Analytics," 2012.

[10] D. Norris, L. Baer, J. Leonard, L. Pugliese, and P. Lefrere, "Action analytics: Measuring and improving performance that matters in higher education," *Educ. Rev.*, vol. 43, no. 1, p. 42, 2008.

[11] "DQC: Data Quality Campaign." [Online]. Available: https://dataqualitycampaign.org/. [Accessed: 01-Apr-2018].

[12] J. P. Campbell, P. B. DeBlois, and D. G. Oblinger, "Academic analytics: A new tool for a new era," *Educ. Rev.*, vol. 42, no. 4, p. 40, 2007.

[13] N. Kondo, M. Okubo, and T. Hatanaka, "Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data," in *Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017*, 2017, pp. 198–201.