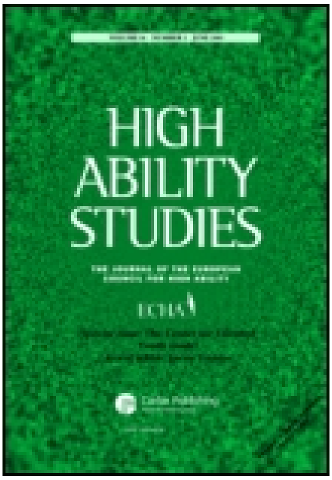

High Ability Studies
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/chas20

Contributions of data mining for psycho-educational research: what self-organizing maps tell us about the well-being of gifted learners

Helena Thuneberg & Risto Hotulainen

Published online: 22 Jan 2007.

High Ability Studies
Vol. 17, No. 1, June 2006, pp. 87–100

Routledge
Taylor & Francis Group

Contributions of data mining for psycho-educational research: what self-organizing maps tell us about the well-being of gifted learners

Helena Thuneberg* and Risto Hotulainen

This article explores applications of the Self-Organizing Maps method (SOM) to psycho-educational data. The study examines the psychological well-being, self-regulatory and motivational styles of pupils at elementary and middle school (N 795). The presentation of the method appears in cases which are related to general education, special needs and giftedness. The aim of this article is to show that SOM provides a unique means with which to visualize, comprehend and interpret psycho-educational data. The SOM method is a convenient method used to identify and study exceptional subgroups and non-linear correlations, as well as to examine theoretical assumptions. The results showed that high academic achievement is related to anxiety, as well as to external and internal pressure, in some gifted subgroups. Such a result is obviously socially constructed and for this reason calls for further study.

Keywords: *Self-organizing maps, Psychological well-being, Gifted*

Introduction

For the researcher, deep knowledge of theory with regard to their numerical data is essential to comprehend the complexity of the collected data and its structure. Although statistical methods can show and prove correlations between studied variables, empirical models which could describe a measure of unity within the data are seldom achieved. Moreover, when the exploration process entails selectively excluding certain factors or variables, the eventual result always represents only a part of the complexity. This has often been argued to be the main shortcoming of empirical studies in the psycho-educational field (cf. Cronbach *et al.*, 1972).

*Corresponding author. Department of Applied Sciences of Education, Faculty of Behavioral Sciences, P.O. Box 26 (Teollisuuskatu 23), FI-00014 University of Helsinki, Finland. Email: Helena.Thuneberg@Helsinki.fi

ISSN 1359-8139 (print)/ISSN 1469-834X (online)/06/010087-14

DOI: 10.1080/13598130600947150

In behavioral science, it is much more realistic to admit that there are often multiple variables which are simultaneously affecting the studied phenomenon and, in consequence, absolute rules concerning studied correlations are seldom achieved. For this reason it would be more than valuable for analyses to utilize the whole range of gathered information, because such an approach could more closely approximate the natural setting. One possibility for this could be the visualization of the data in such a way that clusters and relationships between variables and cases become apparent. By visualization of the numerical data, the researcher is able to notice the existence of assumed theories. Simultaneously, they can observe the appearance of both linear and non-linear correlations between the studied variables. When all the collected information is visible, it is possible to deepen one's understanding of the data's structure as a whole. Additionally, most researchers would be curious to know and observe how their data and constructed variables are intercorrelated according to the main theoretical assumption(s) of the study. Until now such approaches have been quite rare in the psycho-educational field. In this cross-sectional study, the SOM method is used to describe and analyze complex psycho-educational data, while also including proposals for practical applications. Such an application in the frame of giftedness research would contribute improved understanding about the characteristics of gifted subgroups; quantitative studies related to the self-concept of gifted learners have shown that there is no clear evidence for the claims that gifted students were, for example, neither less socially popular nor more intrapersonal than average students, as is often argued (cf. Hoge & Renzulli, 1993). However, qualitative studies have shown that such cases and sub-groups with such characteristics exist among the highly able (cf., e.g., Simonton, 1999). To explain such controversial situations, alternative methods such as SOM can provide additional knowledge which could illuminate the nature of the sub-groups.

The purpose of this article is, first, to provide a theoretical background for SOM, second, to show how SOM can be used to provide added value to the existing data-mining methods in general circulation and, third, to provide a series of simplified examples of how to use SOM on psycho-educational data. Finally, some examples related to the gifted sub-groups from Thuneberg's (2005a) study, which focuses on school motivation and psychological well-being, will be briefly introduced.

Formulation of the SOM networks: a brief theoretical background

The Self-organizing map (SOM), which was created by Finnish scholar Teuvo Kohonen (1988), is based on a new generational understanding of neural networks which process information in a novel way. It is one of the methods which have been said to cope with complex real-world situations in more malleable ways than methods based on traditional mathematical models. Nevertheless, these other methods, such as probabilistic reasoning, fuzzy logic or artificial intelligence, lack features, according to Kohonen, which are characteristic of neural networks. Thus, it enables improved coping in certain situations, for example, when dealing with a great deal of noisy, ill-defined data, non-linearity, signal dynamics or collective effects. One distinctive characteristic is that neural networks can automatically create higher abstractions from raw data. (Kohonen, 1997). As mathematical models, SOM neural networks share similar properties with the statistical multivariate and cluster analyses. When comparing with other data-mining methods based on Bayesian models, SOM creates shaded maps according to the chosen number of the clusters and trainee variable(s). The resulting map is topologically ordered, that is to say: the spatial location of each neuron corresponds to a particular domain or feature of input data.

A special feature of SOM is the way non-linear correlations are derived directly from data. This approach has now become more accessible for all researchers since the development of the increased calculation power of personal computers. The SOM applications have already been tested and used for many years in areas such as artificial intelligence, business, medicine, technology, engineering, speech recognition and so on (Nurkkala, 2004; Honkela, 1997; Kaski, 1997).

Networks are formed as a result of data competition, which begins with a training algorithm. For this purpose there is a need to choose a trainee variable which will give a co-relational dimension for the whole data set. Training has four different phases: first; a reference grid is formed according to the size of the input data sample. In this phase the weighted vectors are placed randomly in every cell of the grid with corresponding colors. Dimensions of the weight are acquired from the input data. In the second phase, input data vectors are placed on the colored dots (nodes) of the grid which share the closest possible weight vectors. When such a match is found, it is referred to as a best-matching unit (BMU). For the third phase, after achieving the BMU, the neighborhood relations begin to affect the forming placements as well. Neighboring prototypes which share similar properties are pulled in the same direction (see Figure 1).

Finally, in the fourth phase, an updating process takes place when final changes are made. In this respect the SOM algorithm is highly iterative. When the training process of the SOM is under way, the data network is 'shrinking' in flexible ways due to data variance. The map itself consists of colored dots. Each colored dot is related to the model which describes the corresponding information unit in question. On the map, the models which share similar information are organized in close proximity to each other. Each information unit is placed on the colored dot which has the best available model for its properties. It could be said that SOM concentrates information by forming a two-dimensional map, which reveals similarities by grouping information. It is, however, worth emphasizing that SOM does not lose any information but, rather, organizes information in a novel way by showing

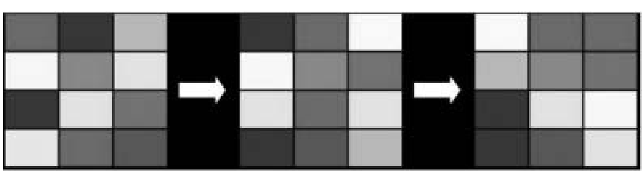

Figure 1. Formation of the map

correlations based on the actual data set. The color-point presentation allows non-linear correlations to become visible.

An important feature of SOM, which is closely related to the training process, is a cluster formation. When using the default option, the number of clusters is formed in accordance with the dimensions of the trainee variable(s). However, a researcher is able to choose the number of clusters, for example, according to the assumed variation of the data. In Figure 2, three different pictures are shown to express the cluster formation process. In this example, data is trained by the grade point average (GPA). Because the GPA was chosen as a trainee variable, all the other information was organized according to this rule. In this particular example, the variable Academic Competence Perception is used.

In Figure 2, on the left the default option has formed three different clusters in accordance with the trainee variable. These clusters indicate subjects with low, middle and high GPA values. The map in the middle reveals through colors where the subjects with high (black), middle (light gray) and low (dark gray) Academic Competence Perception values are located. In the third map, both these phases are combined. The colored clusters indicate that students with low Competence Perceptions are located in the lowest GPA cluster and vice versa. In reality all these phases are formed simultaneously during the training process.

Added value of SOM

Briefly, the main beneficial features of SOM comprise its power to group data efficiently according to a preconceived model and its capacity to visualize the structure of the data set. When viewed from the psycho-educational perspective, such an approach brings observation closer to the natural setting.

The visualization facilitates understanding of the highly dimensional data. The SOM method organizes data according to the properties of the whole data set and then formulates a 'net' of data where information points are interconnected and replaced according to previously created mathematical models. Correlations between the studied variable and the training variable become clearly perceivable by changes in color. When the colors black (low values), light gray (mid values), and dark grey (high values) appear either in this order or vice versa, (cf. Figure 2) it

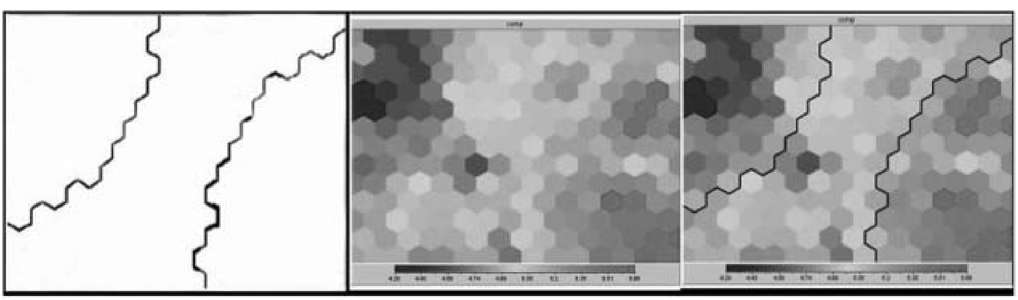

Figure 2. Cluster formation

indicates the existence of a correlation between the trainee variable and the variable in question. In essence, this colorful map can reveal the existence of previously theoretized assumptions and even create new ones.

In addition, a colorful presentation makes it possible to find subgroups, which do not follow the main theoretical assumptions. For example, when blue dots are found in the middle of the yellow or red area, this signals the presence of deviant subgroups. When either blue or red nodes are forming two clearly separately located areas it could be considered as a sign of a non-linear correlation. Both of these exceptional findings are easy to perceive at a glance. From this point of view SOM gives very quick and comprehensive feedback to the researcher. There are also practical tools which allow closer study within the identified subgroups. Simultaneously, the colorful presentation makes it possible to easily detect outliers from the data set. Outliers are easy to rule out by checking the scale under the map (cf. Figure 1).

A common problem with real-world data is the issue of missing values. Some subjects can make human errors and some answers can be late or missing, due to other human reasons. However, with SOM, those observations with missing components can be used as a part of the presentation. The SOM method locates subjects with missing information according to both case and data-set-related information.

In summary, the researcher can simultaneously explore the existence of their theoretical assumptions and may find certain patterns, for example, related to the subgroups, which do not follow the main theoretical assumptions. Such a colorful presentation format enables researchers to readily compare, share and discuss their research findings. By contrast, statistical information does not necessarily leave much space for alternative views and interpretations. SOM facilitates shared information, which can lead to a more comprehensive perception of the phenomena under study. Furthermore, it is worth underlining that, like virtually all data formats, SOM is mainly useful in studies where the numerical data is both valid and reliable.

Illustrating SOM through examples from Thuneberg's study

Before going deeper into the description of the example study, there is a need to emphasize that the study in question is only explained in such depth as is necessary to comprehend the usefulness of SOM. It serves as an example, and not as an independent research report. Persons who are interested in the contents and results of the study should see Thuneberg, 2005a; or in Finnish: Thuneberg, 2005b; Thuneberg, 2005c.

Description of the study

This research, described as a 'Psychological well-being at school project,' consists of several sub-studies which explored the self-regulatory behavior, motivational styles and psychological well-being of pupils in elementary and middle school. For the SOM examples, the following four sub-studies were chosen and described more clearly.

The first sub-study selected for this SOM-application example is based on Wade and Moore's study 'Experiencing special education' (1993). This study obtained students' views about several school-related areas, ranging from lessons and teachers to feelings of being different and parents' role at school. The other three studies used the Academic Self-Regulation and Pro-Social Self-Regulation questionnaires and the Basic Needs Satisfaction at School (Work) questionnaire which were based on the Self-Determination Theory (SDT) of Deci and Ryan (2002). The questions in these scales were targeted to discover, for example, how children's environments manage to fulfill the following three basic needs: competence, relatedness and autonomy. Several studies have shown that fulfillment of these needs is in jeopardy, especially if a pupil has learning difficulties or other problems (Deci *et al.*, 1992). Environments which do not respond to pupils' needs can lead students to a state known as 'learned helplessness' (cf. Weiner, 1993). According to the SDT, psychological well-being is not dependent on the achievement level of the child. For this reason, a high capability in one area, or overall giftedness does not, of itself, guarantee a person's sense of well-being (cf. Dweck, 1999; Hotulainen, 2003). Not every high-achieving or gifted child automatically feels autonomous, related or competent. To find out how various environments affect students' school experiences and feelings, the study focused not only on pupils from special education classes, but also on pupils from highly selective classes, Music classes (test criterion musical ability) and English classes (test criteria fluent English and general verbal ability).

Variables. The independent variables of the study were gender, general versus special and selected education, age, academic achievement (GPA). Dependent variables of the study were formed according to the SDT and related questionnaires. In Table 1 the psychometric properties of the Finnish versions are presented.

Reliability estimates of the scales were mostly satisfactory (Cronbach's over .70). However, the variables 'Autonomy', and 'Competence' from the BPNS, and the variables 'Not feeling different' and 'Parents and school' from Wade and Moore, showed lower values. Consequently, any interpretation of the results using these variables should be drawn carefully.

Subjects. The subjects in this study were selected both from elementary and middle schools. The sampling method could be called 'a non-proportional quota sample' (Trochim, 2005), because the aim was to have as many special education groups as possible included. All sixth and ninth-grade special education classes of the city of Espoo (200,000 inhabitants) participated. In addition there were general-education classes and selective classes (from grades three to nine). A sample of 150 pupils was from Joensuu, a small town in eastern Finland. The selective classes included Music classes and English classes. The total number of the participants was 796, of which 55% were boys. Mean age was 12 years and three months.

Table 1. Descriptives and reliabilities of the dependent variables of the study

	Items	Mean	SD	Alpha	Total
SRQ-A					
External (exte)	9	2.61	.60	.79	732
Introjected (intro)	9	2.56	.61	.84	732
Identified (ident)	7	3.11	.56	.81	732
Intrinsic (intri)	7	2.32	.66	.84	732
SRQ-P					
External (extb)	5	2.62	.58	.71	731
Introjected (intob)	10	2.99	.65	.84	731
Identified (ideb)	10	3.43	.54	.89	731
BPNS					
Autonomy (bpnsauto)	6	4.69	.96	.59	735
Competence (comp)	6	4.96	.89	.57	735
Relatedness (rela)	8	5.07	.96	.76	735
WM					
Not getting into trouble	6	4.10	.58	.74	760
Friends and enemies	6	4.03	.51	.67	750
Not feeling different	5	3.83	.53	.54	753
Teachers	6	3.78	.63	.72	767
Lessons	6	3.57	.57	.65	756
Being assessed and tested	6	3.48	.72	.75	765
Parents and school	4	3.41	.63	.58	764

The SOM examples

Example 1: The whole data set presented simultaneously. The iteration process produced a picture in which the whole example data is shown at the same time by a set of colored maps (Figure 3). Each map presents a variable and shows its correlation to the trainee variable and its clusters — in this example of the three GPA clusters. The maps can be studied either in the input order or in (a) correlation order. The latter means that variables which have strongest correlation are shown as neighbors.

By using SOM, one is able to evaluate the overall qualities and structure of the well-being and motivation data in a more comprehensive way. Such a comprehensive view becomes possible because all the variables can be explored and compared simultaneously. At the same time, SOM allows an examination of the following: first, variance within the cluster (color tone); second, variance between the clusters within the variable in question; and finally, variation between all study variables.

Example 2: Using SOM in probing theoretical assumptions. The Self-Determination Theory argues that motivation changes from extrinsic towards intrinsic along a continuum. According to the theory, the most extrinsic variable, 'External,' correlates most closely with its neighbor, 'Introjected,' and less closely with the following 'Identified,' and the least with the most intrinsic variable (Deci & Ryan,

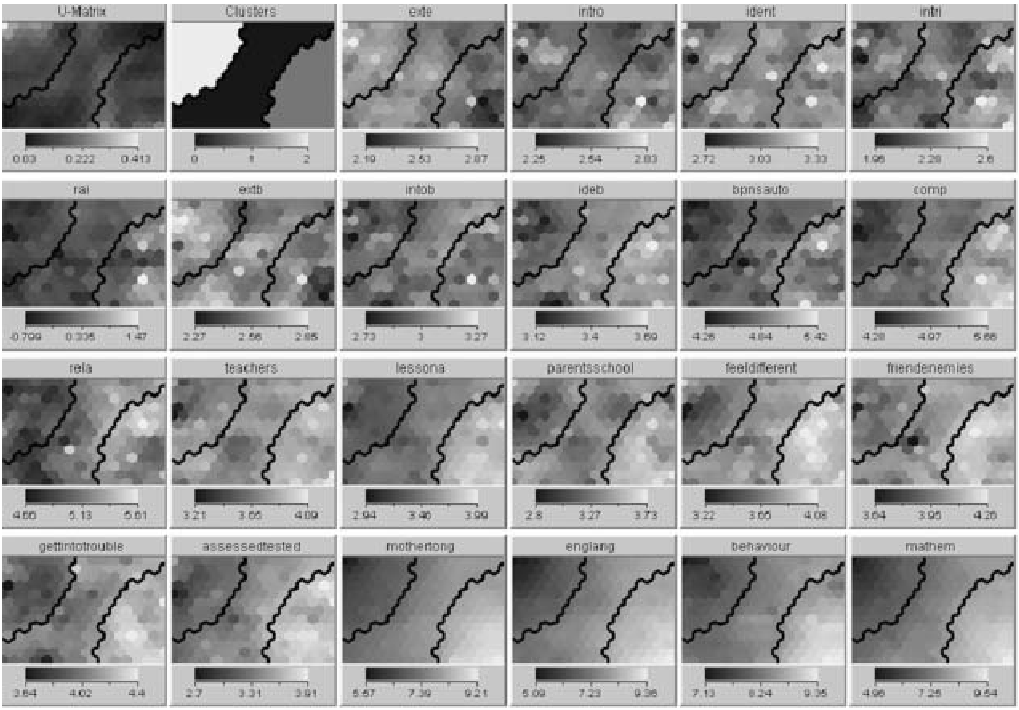

Figure 3. All variables of the study are shown as SOM maps

2000). Although it is possible to study these claims by statistical correlative analysis, we also used SOM. Here, the data have been trained by the Relative Autonomy Index, RAI. The maps confirmed the existence of a theoretical motivational continuum of the SDT, in that it was relatively simple to see how the colors, step by step, map by map, changed towards their opposites. Finally, the map of the RAI showed the positive correlation with the more intrinsic variables (identified and intrinsic) and a negative correlation with the more external variables (external, introjected), as Figure 4 illustrates.

Example 3: Using SOM in identifying exceptional groups. The nodes are the small particles representing data points which then form a map. The most similar students share the same node or at least they are placed in nodes which are located close to each other. For this reason, the neighbors usually have very similar information (although not necessarily in relation to every variable). All the nodes inside a cluster share the same values in relation to the trainee variable, in this case the GPA. By using SOM, one is able to readily recognize the possible outliers and exceptional groups since they are the nodes which deviate from their surroundings by a different color or color tone. Thus, the SOM makes it possible not only to compare the correlations, but also to explore the non-linear connections. Figure 5 shows in the variable ‘Friends and enemies’ a few nodes, which are black or dark grey, while the whole neighborhood is light grey. This particular area describes students’ experiences

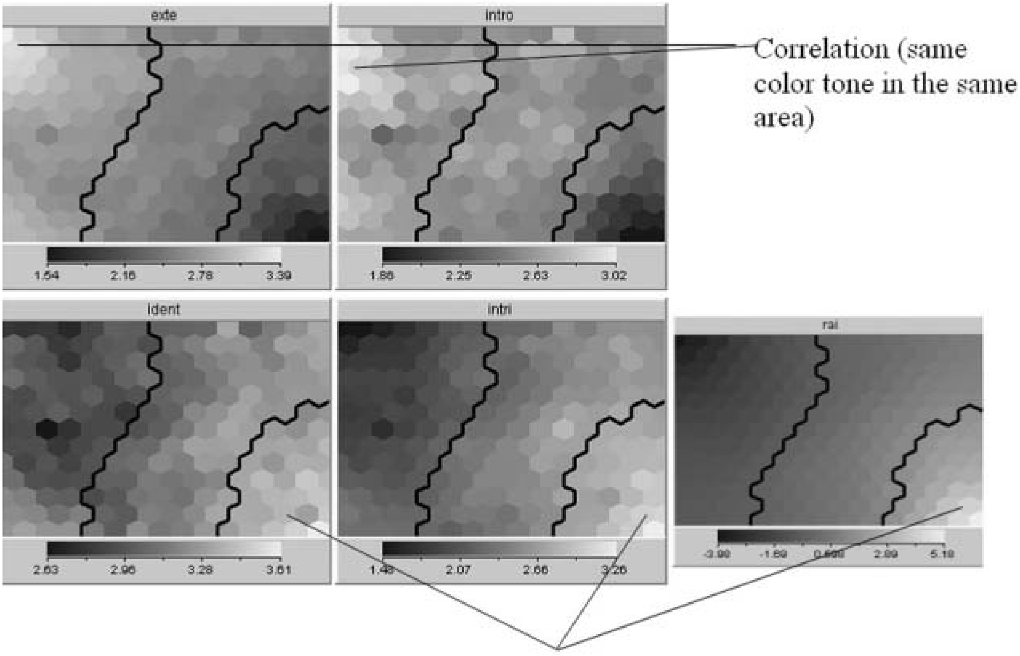

Figure 4. Correlations between motivational variables and a trainee variable Relative Autonomy Index, RAI

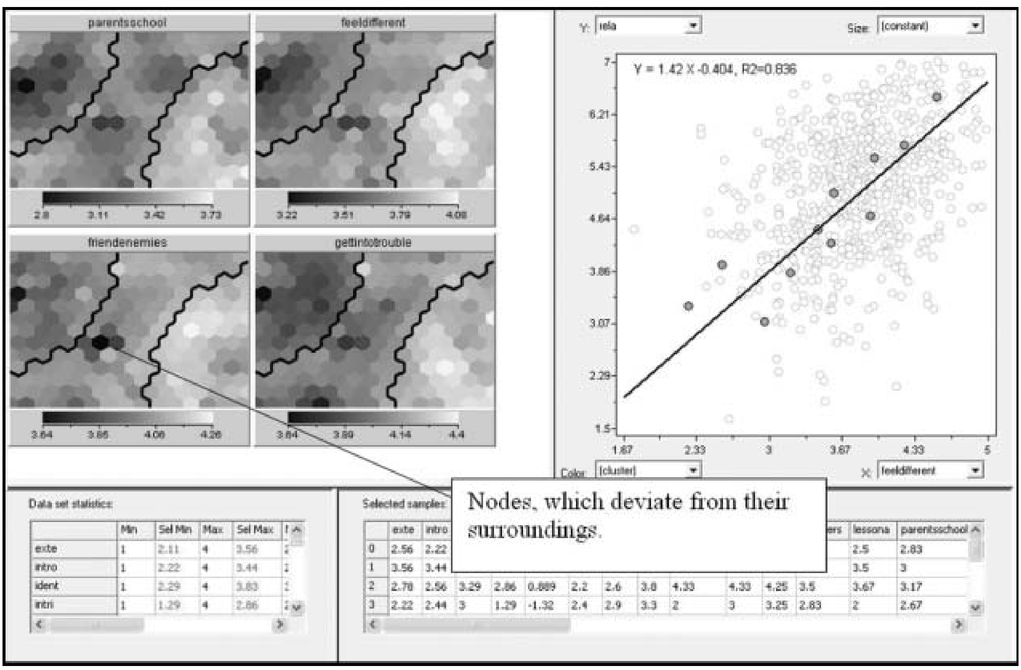

Figure 5. Showing selected exceptional groups in four windows

with friends. These exceptional groups have had terrible or quite bad experiences. According to previous knowledge (Lane, Wehby & Barton-Arwood, 2005), it could be assumed that they were more likely to be found in the lowest GPA cluster.

As a researcher, school principal or psychologist, you might now be interested in studying which students belong to each of these groups, and how these groups relate to the other variables. The exploration of the maps could be supplemented by inspecting the scatter-plots, statistical descriptives, and raw values of the students. Note that the same subjects in every map are situated in the same node.

Example 4: Using SOM in studying different school related dimensions. The fourth theme brings together information about various school related dimensions, such as task-orientation, anxiety, and attitudes towards tests. By using SOM, these dimensions can be explored in a concrete, visible manner.

In the first case, the data was trained with the 'Identified' variable, which is one of the middle variables on the motivational continuum (Figure 6). The Identified

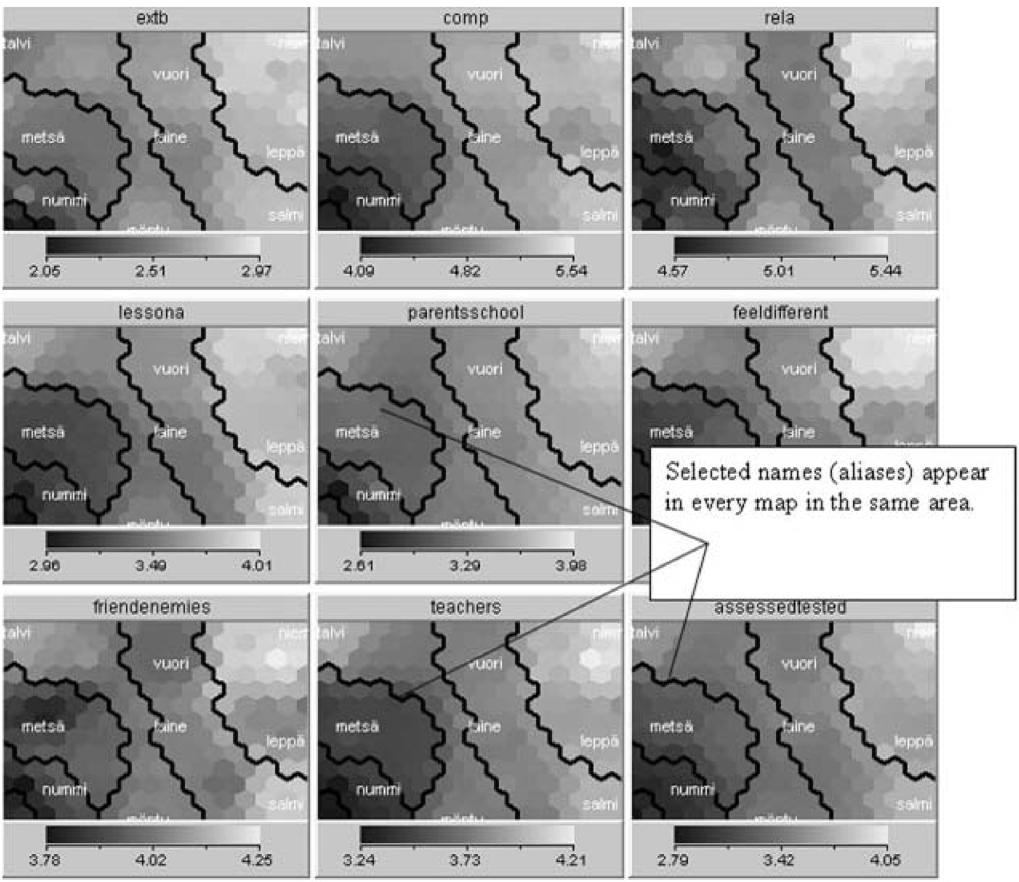

Figure 6. Identification: enabling the possiblity to find out how each student [aliases used for research ethical reasons] behaves in relation to the variables

variable is considered to indicate school satisfaction, good adjustment at school, and acceptance of school goals (Grolnick & Ryan, 1989; Ryan & Connell, 1989). Because it was considered interesting to study the extreme ends of school satisfaction, the number of clusters chosen was six, in that this made it possible to study and identify those students having the worst or best feelings about school, and then follow how this experience related to, for example, the parents' role at school, lessons, or teachers.

Additionally, it is possible to identify students by checking the raw values of the selected cluster. It is possible to find out, for example, which of those students had positive experiences with teachers, what the level of school achievement was, and any other information that was useful and available. There also exists the alternative to pre-select those students of interest. This option, entitled 'label picking' shows names (aliases) and their location on the maps (Figure 6).

In the second case, we wanted to compare a highly selected, gifted group (Music class, the three best GPA clusters out of eight clusters) with the others. We used the aforementioned methods. By using the maps, it was easy to pick up the groups, cluster by cluster, study the scatter-plots and regression-lines, and then test the significance of the correlation coefficients' differences by other methods. By combining SOM and significance testing, we discovered that there was an interesting subgroup of girls. In the selected girls' ($n=22$) group, the Introjected variable was much more strongly related to the GPA than in the others' ($n=552$) group. When statistically comparing these, the correlation coefficient difference turned out to be significant ($z=3.39$, $p<.001$). Further comparisons showed, in addition, that the external regulation variable was more and positively related to the GPA in the selected girls' group, but negatively in the others' group. Again the difference was significant ($z=2.6$, $p<.001$).

These two findings indicated that the highly selected girls of the Music classes experienced anxiety in relation to achievement. The Introjected self-regulation is shown to be connected with outside (from significant adults) and inside (from self) pressure (Grolnick & Ryan, 1989; Ryan & Deci, 2000). The connection between the external self-regulation and the GPA shows that these girls seem to study more, because they want to gain incentives or to avoid punishments more than the others (Ryan & Deci, 2000). They study because of the grades but not because of the task itself. The formation of such orientation is obviously socially constructed and for this reason calls for further studies (cf. Heller & Ziegler, 1996). If this or another especially gifted group, or another certain group, was a target of interest, the self-organizing maps could be executed only for this group. This would tell whether there are notable differences inside the group.

Thus, by using SOM we were able to evaluate the students' situation of overall psychological well-being, we could study groups which behaved in a manner which was not as hypothesized, and could become aware of how specific students, who were of concern beforehand, experienced various school related areas. Thus, we had a good practical basis on which to formulate plans, interventions and further research.

Conclusions

Self-organizing maps illustrate the features of ordinary psycho-educational data, such as correlations and directions, in a readily observable manner. The method is useful in making the comparisons between the variables easier, faster and more obvious than by any other means.

By use of the maps which are generated, it is relatively easy to recognize the existence of the theoretical models and their linear correlations by the various colors. An additional benefit comes from the fact that visualization can simultaneously reveal non-linear correlations between different factors which might not otherwise be predicted but which affect the data. In this sense SOM can be used to test theoretical assumptions and to create new ones. In the psycho-educational field, SOM can show common and expected features among average subjects, but it can also reveal hidden characteristics of the sub-populations (e.g., underachievers, gifted, depressive children, etc.), which do not share the same properties as the majority of the population. For the researcher, such an approach gives valuable information about their data. The possibility of electing and filtering desired groups and examining them further in other windows (scatter plots, statistical descriptives and exact/raw value windows) allows deeper and more detailed analysis. For example, it makes it possible to examine both the common trends and the deviations: groups which act as they were hypothesized to act, and groups which surprise us. The maps provide the researcher an opportunity to manipulate the settings of the data from various points of view, in order to see the study's many dimensions. The examples of SOM that have been used here have shown that it is also a valuable and reliable tool for the identification of a special group.

Additionally, SOM can have practical value for the school authorities such as principals, school psychologists, school counselors, and special education teachers. With SOM one is able to study the school, the class, and specific student related variables. It can offer an improved understanding of the factors leading to successful learning, and through its process of visualization, communication between groups such as teachers and policymakers is enhanced (Hautamäki & Nurkkala, 2004). As an educational tool, it can reveal relations between those learning-related factors which are typical characteristics for a particular population or, alternatively, for a particular environment. By using SOM, one is able to identify the students at risk of maladjustment and even the environmental antecedents of such behavior. This enables both the teacher and administrator the chance to plan educational interventions and environmental changes. For instance, the concept of under-achievement and its relation to giftedness and its appearance in the school environment can be examined in relation to motivational variables. If high academic achievement is a consequence of anxiety as well as external and internal pressure in some subgroups, as the SOM maps indicate, we have to study those variables which feed into such orientations more closely.

Although the SOM method is unique in many respects, it can also be used as a supplement to other statistical methods. For a comprehensive analysis, these other statistical methods are necessary, beginning from significance testing by analysis of variance to more sophisticated multivariate analysis methods. In summary, however, there is a clear need for tools like self-organizing maps, because they can provide unique concreteness in abstract matters. As discovered in this study, SOM provided valuable information about the gifted learners; also high-achieving students can be at risk for psychological ill-being and the basis of their motivational orientation can be unfavorable for their learning in the long run. Such information can enhance the multidimensional understanding of psycho-educational phenomena such as giftedness, which is surely a significant goal in the whole research process.

References

Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972) *The dependability of behavioral measurements: theory of generalizability for scores and profiles* (New York, John Wiley).

Deci, E., Hodges, R., Pierson, L. & Tomassone, J. (1992) Motivationally relevant predictors of achievement and adjustment in learning disabled and emotionally handicapped students, *Journal of Learning Disabilities*, 25(1), 457–71.

Deci, E. & Ryan, R. (2000) The 'what' and 'why' of goal pursuits: human needs and the self-determination of behavior, *Psychological Inquiry*, 11(1), 227.

Deci, E. & Ryan, R. (2002) *Handbook of self-determination research* (Rochester, NY, University of Rochester Press).

Dweck, C. S. (1999) *Self-theories: their role in motivation, personality, and development* (Philadelphia, Taylor & Francis).

Grolnick, W. & Ryan, R. (1989) Parent styles associated with children's self-regulation and competence in school, *Journal of Educational Psychology*, 81(1), 143–54.

Hautamäki, J. & Nurkkala, T. (2004) *Learning to learn assessment data represented in self organizing maps.* Conference presentation presented 1 December 2004 at the Online Educa Berlin. Available online at: http://www.dipoli.tkk.fi/europro/matkat/oeb2004/workshops.html (accessed 5 May 2005).

Heller, K. & Ziegler, A. (1996) Gender differences in mathematics and the sciences: can attributional retraining improve the performance of gifted females? *Gifted Child Quarterly*, 40(4), 200–211.

Hoge, R. D. & Renzulli, J. S. (1993) Exploring the link between giftedness and self-concept, *Review of Educational Research*, 63(4), 449–65.

Honkela, T. (1997) Self-organizing maps in natural language processing, Espoo. Available online at: http://www.cis.hut.fi/~tho/thesis/index.html (accessed 13 October 2005).

Hotulainen, R. (2003) Does the cream always rise to the top? *Correlations between pre-school academic giftedness and perceptions of self, academic performance and career goals, after ten years of Finnish comprehensive schooling* (University of Joensuu, Publications in Education), 84.

Kaski, S. (1997) Data exploration using self-organizing maps, *Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series* (Finnish Academy of Technology), 82.

Kohonen, T. (1988) *Self-organization and associative memory* (New York, Springer-Verlag).

Kohonen, T. (1997) *Self-organizing maps* (New York, Springer-Verlag).

Lane, K., Wehby, J. & Barton-Arwood, S. (2005) Students with and at risk for emotional and behavioral disorders: meeting their social and academic needs, *Preventing School Failure*, 49(2), 6–9.

Ryan, R. & Connell, J. (1989) Perceived locus of causality and internalization: examining reasons for acting in two domains, *Journal of Personality and Social Psychology*, 57(1), 749–61.

Ryan, R. & Deci, E. (2000) Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being, *American Psychologist*, 55(1), 68–78.

Simonton, D. K. (1999) Significant samples: the psychological study of eminent individuals, *Psychological Methods*, 4(4), 425–51.

Thuneberg, H. (2005a) *Self-determination, motivation and the fulfillment of the psychological needs at school.* Conference paper presented 3 August at the ISEC2005, organized by University of Strathclyde, Scotland.

Thuneberg, H. (2005b) Academic self-regulation at school, in: S. Havu-Nuutinen & M. Heiskanen (Eds) *Integrating and differentiating paths of learning and education* (Faculty of Education, University of Joensuu. Joensuu, Finnish Educational Research Association), 76–89.

Thuneberg, H. (2005c) Psychological well-being in kindergarten, in: E. Kontu & E. Suhonen (Eds) *Special pedagogy and early childhood* (Helsinki, University Press), 95–108.

Trochim, W. (2005) The research methods knowledge base (2nd edn). Available online at: http://trochim.human.cornell.edu/kb/index.htm (accessed 16 January 2005).

Wade, B. & Moore, M. (1993) *Experiencing special education* (Buckingham/Philadelphia, Open University Press).

Weiner, B. (1993) On sin versus sickness, a theory of perceived responsibility and social motivation, *American Psychologist*, 48(1), 957–70.