Graphical artefacts: Taxonomy of students' response to test items

Oduor Olande

Published online: 28 June 2013

© Springer Science+Business Media Dordrecht 2013

Abstract The present study, carried out in the Nordic countries, examines the characteristics of students' scholastic performance on items containing graphical artefacts, that is, bar graphs, pie charts and line graphs, selected from the Programme for International Student Assessment (PISA) survey test. Graphical analysis of statistical data resulted in the observation of two major categories of performance by the students. The results of cluster analysis also confirmed the two approaches. One approach consists of items perceived as requiring identification, that is, focusing primarily on perceptual elements. The other consisting of items requiring a critical-analytical approach, that is, involving evaluation of the graphical system, active interaction with subject specific operators and forms of expression. The general observation is that the pattern of response is similar for all these countries, with items demanding an identification approach showing comparatively higher scores than for items perceived as demanding a critical-analytical approach.

Keywords Cluster analysis · Graph comprehension · Graphical artefacts · Mathematics literacy · Semiotics

1 Introduction

Graphical artefacts are important in the mathematics and science teaching-learning [*didaktik*] process since they can be used to advance the understanding of scientific and mathematical objects as well as functioning as entities for fostering what is in the present study, perceived as critical-analytical and communication skills (Balchin & Coleman, 1965; Bertin, 1983/1967). Graphical artefacts are not only handy entities in scientific analysis, they are also central to the rhetoric of scientific communication (Roth, 2003). It is perhaps the communication dimension exemplified by the simplification and summarizing of concepts and phenomenon that informs the frequent use of graphical artefacts in societal communication for example in advertisement and news supplements. It is alleged that graphical

O. Olande (✉)

Department of Applied Science and Design, Mittuniversitetet, Holmgatan 10, 851 70 Sundsvall, Sweden
e-mail: oduor.olande@miun.se

artefacts are probably the best way to communicate complex technical information to an intelligent, lay audience (Wainer, 2009). While making sense of graphical artefacts may seem straightforward, this might not necessarily be the case and as such one is expected to be conversant with the convention associated with handling, and to some extent the knowledge of the context, phenomenon or object to which the graphical artefact relates (Curcio, 1987; Lowrie & Diezmann, 2007).

The present study examines the results on assessment items containing graphical artefacts from students in the Nordic countries. Most of these students were about 15 years of age, and in the final year of compulsory school (cf. Organisation for Economic Co-operation and Development—OECD, 2007). The Nordic countries are included in the study based not only on their cultural, educational and geographical proximity, but also on the fact that other studies have indicated the presence of similar patterns of performance for these countries in standardized tests in both science and mathematics (Kjærnsli & Lie, 2004; Lie & Roe, 2003; Olsen, 2006; Olsen & Grønmo, 2006). This observation is also significant from a statistical data analysis and sample selection perspective. The significance of using PISA test survey emanates from the aim of PISA, which is to test the preparedness of students for adult life. The premise in the present study is that the adult life alluded to here is increasingly characterized with graphical communication since graphical competence is vital for effective citizenship. Also the data material from PISA is perceived as unique since, unlike some standardized tests for example the Swedish Scholastic Aptitude Test (SweSAT), there is no designed section specifically testing students' graphical abilities. Thus, in this case students' interaction with graphical artefacts may be considered as occurring in 'authentic' situations that is, where the focus is not entirely on a casual interaction with the graphical artefact but also on the application of other mathematical tools and forms of expression.

1.1 Rationale

In spite of graphical artefacts having been a part of human prehistory and given the ease with which these can be produced and accessed through various ICT technologies and media, the skills and abilities needed to make sense of the same seem to evade our understanding. According to Roth (2009), despite the interest and commitment to graphs and graphing, what it takes to interpret a graph is much less clear. Thus, it is reasonable to assert that interacting with graphical artefacts is a complex undertaking involving more than just visual dimensions (Åberg-Bengtsson & Ottosson, 2006; Radford, 2009; Raschke & Steinbart, 2008; Winn, 1993). Some of these aspects are considered to be cognitive in nature and thus some level of understanding these cognitive components is crucial for the effective interaction with graphical artefacts (Radford, 2009; Raschke & Steinbart, 2008; Winn, 1993). From a mathematics *didaktik* perspective, graphics are artefacts that might be transformed by or might require the application of other mathematical tools or symbols in the sense making process. The significance of this observation is realized considering that "[s]ymbols are indispensable as means for coding the result of thinking. More importantly however, symbols also function as means for the regulation of the thinking process" (van Oers, 2000 p. 136; see also Vygotsky, 1978). Thus, the extent to which students are able to manifest effective use of mathematical symbols (operators and forms of expression) in making sense of graphical artefacts might be perceived as an indicator of their 'competence' in using these tools thus reflecting how they actively and imaginatively (creatively) endow the conceptual objects encountered in their culture with meaning in the given context (Radford, 2008).

From a Vygotskian research approach, tools [artefacts] are perceived as playing an important role in the cognitive process (see Radford, 2003; Vygotsky, 1978). The use of signs leads humans to structures of behaviour that break away from biological development and creates new forms of a culturally-based psychological process (Vygotsky, 1978). The constructs tools and signs in the present study are not restricted to material tools but are regarded as encompassing semiotic means of objectification such as language, writing, algebraic symbol system, diagrams, maps etc. (Radford, 2008; see also Daniels, 2008). Wertsch (1985) also suggests that Vygotsky's theoretical approach can be understood by recognizing its three interrelated general themes of which the theme he posits as analytical prior to the others is "the claim that an adequate account of human mental functioning must be grounded in an analysis of the tools and the signs that mediate it" (1985, p. 15). In the present study an analysis of students' response to assessment items is done with respect to the application of subject specific tools and forms of expressions. This is based on the notion that the investigation of students' interaction and use of semiotic means of objectification is a methodological way of accounting for learning (Radford, 2008). However, it ought to be noted that 'tools' are not per se, neutral: it is not uncommon for a hammer in the hand to be associated with the *likelihood* of being used to exert directed force on another object, for example, hitting a nail (thus it is not unusual to associate a hammer with a nail). Consequently, a sign or symbol cannot be considered in isolation because it is always bound up with intentions, motives and the object of actions (Roth, 2008).

Assessment practice in the school system exemplifies an activity that is used to appraise the 'competency' of students to use subject specific tools and forms of expression. In this regard it is expected that one of the major purposes of assessment is to ascertain the nature and quality of the responses to test items and hence their mathematical thinking. Based on empirical research on students' response to test questions, Bergqvist, Lithner and Sumpter (2003) observed that most students have difficulties interacting with items whose demands are perceived as going beyond the use of simple procedural operations (see also Goodchild & Grevholm, 2007). This can partly be explained by the nature of the mathematics, that is, the use of symbols and concepts that are withdrawn from or might be divergent with everyday vocabulary (Österholm, 2008; Roe & Taube, 2006; Sfard, 2001). It is this very nature of mathematics characterized by demand for precision that carries with it a rigidity that may not augur well for the training of critical ability (Balchin & Coleman, 1965). Following these arguments, it is probable that 'mathematics' tasks per se might not necessarily be a better indicator of the critical-analytical and creative stance of students. In view of what may be considered as a limitation of mathematics, Balchin and Coleman (1965.) put a case for the elevation of the use of graphics in teaching and learning situations. A similar case for graphical representation is made by Bertin (1983) who underscored the importance of graphics in pedagogy since they can introduce the basics of logic and the essential processes of analysis and decision making into all disciplines and might help in revealing the intelligence of weak students. Thus, in the quest to study interaction with graphical artefacts it is worthwhile to monitor the performance of students in test items with graphical elements.

2 Making sense of graphical artefacts

As mentioned earlier, making sense of graphical artefacts is a multifaceted and complicated process (see e.g., Lowrie, Diezmann & Logan, 2011; Radford, 2009; Raschke & Steinbart, 2008). Thus, it is given that in an assessment environment, students might have to deal with

more than a casual interaction with data. In a quantitative study based on a test environment, Gustafsson, Wedman and Westerlund (1992) observed that a subtest on the interpretation of diagrams, tables and maps (DTM) closely fitted or coincided with the descriptions of the overall factor (what they referred to as the G-factor) that reflected the capacity for problem solving in new and complex tasks. Following up on an earlier study where a quantitative factor was isolated from the subtest DTM, Åberg-Bengtsson (2005) points out that the DTM subset was assumed to measure analytic dimension and that the quantitative factor in DTM was related to items where calculations of one kind or another were used in addition to the reading off of values. Guthrie, Weber and Kimmerly (1993) conducted a study involving local *search* and *global search*. Local search was perceived as involving low level questions dealing with extraction of information, that is, the search for specific information from the graphical artefacts, while global search involved some level of ‘abstraction’. In global search one is expected to deal with complex sets of spatial, temporal, and causal relationships, identify relationships contained in the illustration and forming a mental model that will function for recall and problem solving (Guthrie et al., 1993). Framed thus, global search is perceived as including what can be considered as *multireferencing*; that is, holding different bits of information or impressions in interaction with a graphical artefact. Ben-Zvi and Arcavi (2001) suggested the concept of global and local perspective to data analysis. Global view is perceived as the ability to search for, recognize, describe and explain general patterns in the data set by general inspection or the use of other parameters or techniques, while a local view of data focuses on single points or entries within a group of data. The base line for global and local perspectives to data is to a large extent also determined by the question at hand. From the research findings above, it seems that there are two dimensions to interaction with graphical artefacts: on the one hand is a dimension that can be referred to as ‘critical-analytical’ and on the other hand there is a visual or ‘identification’ dimension.

Bertin (1983) stated that the visual dimension (perception) consists of defining relationships established within or among the image(s) or between the image and the real world; as such he posits that the maximum utilization of the perceptual powers within the framework of logic reasoning is the true purpose of graphics. His emphasis on the power of visual perception can be understood in the light of pioneers of graphics such as William Playfair and Charles Joseph Minard who intended to make it easier to visually interact with graphical artefact (see Costiagn-Eaves & Macdonald-Ross, 1990; Funkhouser, 1937; Robinson, 1967). Inasmuch as visual perception might not necessarily account for other factors such as familiarity with context, it is through it that areas of interest in the graphical artefacts are identified. Based on experience and familiarity, some relation or connection is made which then leads to appropriate actions and interaction such as, transformation, manipulation, translation or reporting. Thus, viewed from the perspective of interaction with a graphical artefacts, ‘identification’ is the initial contact with a graphical artefact and when no further *active* action such as manipulation, translation, transformation is executed, the results of the interaction can be said to be ‘self evident’. However, it is not uncommon that some kind of *active* action might be called for after ‘identification’. It might be necessary to critically analyse the contents, the need to employ subject specific tools as operators, and then make some pronunciation, in other words, communicate the solutions or findings etc., thus being ‘critical-analytical’. Roth and Bowen (2001) identify two mutually constitutive processes in use while interacting with graphical artefact, namely *structuring* and *grounding*. Structuring entails identifying relevant signs and tools, mostly by visual inspection that might be used for effective interaction with the graphical artefact. Grounding, on the other hand, involves the linking of an inscription (a graphical artefact) with a network of referring relationships in the meaning making process. Clearly, both ‘grounding’ and being critical-analytical rely very much on experience; experience with the

demands imposed by the graph, that is, familiarity with the conventions attached to the graphical system at hand, identifying the appropriate signs and tools e.g. the need to apply mathematical concepts or operators and forms of expression in interacting with the graphical artefact. To a lay person not well versed with a stem and leaf diagram for example, it can be intimidating and outright difficult, whereas the same may not apply to a student of statistics, a statistician or a child who has been exposed to an environment where there is frequent use of the diagram and to whom these aspects may be self evident.

In a study on graphic information learning by Postigo and Pozo (2004), it was found that two factors had a positive effect on graph learning performance, namely level of education and subject knowledge of domain questions. Using the Graphical Languages in Mathematics Test (GLIM), Lowrie and Diezmann (2007) were able to show that the difference in success on graphical items can be attributed not only to the knowledge of the embedded mathematics content in the question but also on the command of the graphical languages, that is, the student's ability to decode information onto and from the graphic. This implies that the achievement on an assessment item involves the effective integration of mathematical procedures and forms of expressions as well as the understanding of the rules and conventions associated with making sense of the different graphical systems. In the present study the assumption is that at the time of taking the PISA test, the students had been exposed to the basic prerequisites for handling items containing graphical elements. This assumption is necessary given that the data availed by PISA does not include actual students' solutions. However, it is also known that some students, even after 8 years of schooling, have difficulties with some of these competencies, for example, ratio, proportionality, the concept of probability (Adjage & Pluvineau, 2007; Jones, 2005). On the other hand elementary arithmetic procedures involving operations such as addition and subtraction is perceived by most students as almost self evident. Thus, it is presumed that if the initial interaction with a graphical artefact (identification) is followed by the elementary arithmetic procedures, then the interaction can also be perceived as involving an identification approach (i.e. it is at a level that is self evident). A critical-analytical approach would involve such methods that go beyond a casual interaction with a graphical artefact.

2.1 Categorizing graphicacy

A number of studies on graphic comprehension have identified different levels of graphicacy, that is, interaction with the aim of making sense of graphical artefacts due to oneself or others. In a review of graphical comprehension conducted by Friel, Curcio and Bright (2001), it is apparent that three levels of interacting with graphical artefacts emerge. Friel et al. (2001) present a study in which the different variants of these levels are outlined. The categories are (i) an elementary level where students are able to identify and read off values i.e. extracting data from the graph; (ii) an intermediate level involving working within the graphical system—interpolating and finding relationships in the data presented in the graph and (iii) an advanced stage involving extrapolating, predicting, or inferring from the graphical system—at this level deep understanding of the structure of the data is perceived as essential. Regarding the three levels of graphical comprehension, Bertin (1983) suggested that one does not *read* a graphic, one asks three questions of it, namely: (i) what are the x and y components of the data table—in other words the structural component of the graph (ii) what are the groups in the x and y that are constructed by the data and lastly regarding the 'information' being displayed (iii) what are the exceptions to these groups. Whereas there is general consensus in research for the elementary level of graphical comprehension, the same cannot be claimed for the intermediate level of graphical comprehension. For example, while Curcio, (1987) suggests that the intermediate level involves interpretation and integration of the information presented in the graph, Bertin (1983) and Wainer (1992) seem to

suggest that this is limited to a part of the graph. This observation points to the drawbacks inherent in what can be perceived as a linear hierarchical construct of cognition.

The three questions and the categories outlined above (see Bertin, 1983; Friel, et al., 2001) seem to share some similarity (Curcio, 1987; Wainer, 1992) and can without loss of generality, be summarized by on the one hand, a set of concerns dealing with visual perception and on the other hand, concerns about data reduction at different levels. However, as it was mentioned earlier, other researchers seem to adopt a general perspective of interaction with graphical artefacts involving two major categories (e.g. Ben-Zvi & Arcavi, 2001; Guthrie et al., 1993). Gal (1998) captured the essence of questioning relating to interaction with (including interpretation of) graphical artefacts by stating that there are principally two types of questions that can be posed: *literal reading questions* and *opinion reading questions*. In literal reading questions students simply have to point to a data set in a graph or table, while opinion reading questions elicits students' ideas about the meaning of the overall pattern of data. Thus, Gal's sets of questions seem to suggest that interaction with graphical artefacts can be focused on visual perception where the 'representation' is taken at face value or where the interaction calls for more a 'mental' engagement involving reflection and analysis.

From the research outlined above it is apparent that the types of questions posed in relation to interaction with graphical artefacts are an important factor in making sense of graphical artefacts. In the context of the present study, the importance of questioning is appreciated inasmuch as these have a potential of informing the choice of action (tools and forms of expression) to be employed in the sense making process. Thus, the present study pursued the categorization of graphical comprehension that pays attention to subject specific tools and forms of expression in interaction with graphical artefacts.

3 The study

In the present study, items containing graphics were selected from PISA 2003 survey mathematics literacy domain. The criterion for selection was that the items were connected to one of the graphical artefacts presenting different ways of representing data; line, pictorial, bar graph and pie-chart. In the present study the items are identified using the original PISA codes. The reasons are twofold: firstly, it makes it easy to identify the items in other PISA related literature; the other reason is based on the fact that some of the items are not yet released to the public domain, a fact that may be compromised in the event that the items are otherwise identified.

The data consisted of the percentage of students registering correct answers in the different countries (p -values). Since these percentages contained information on the overall achievement for the countries and the overall difficulty of the items (Olsen, 2005)—this being information deemed to be sufficient for the study—it was decided not to transform them further. Also transforming the data further by say, standardization, may or may not be feasible and sometimes has the effect of dampening the data (Kaufman & Rousseeuw, 2005; see also Dudaite & Eljio, 2008, September). A total of 19 items were selected. These are shown in Table 1 based on PISA classification framework, showing the item format and their competency clusters. (Sáenz, 2008, has also referred to these competency clusters as level of theoretical complexity.)

The clusters of competencies are based on the kind of cognitive demand or process through which test items are encountered. These can be perceived as the cognitive processes which should be activated to connect the real world in which the problem arises with the mathematics and then to solve the problem posed (OECD, 2003; Sáenz, 2008). In particular *reproduction cluster* of competency, involves the reproduction of known knowledge, performance of routine procedures, application of standard algorithm manipulation of expressions containing symbols

Table 1 Selected items classified using aspects of PISA framework

Item	Item format	Competency
M302Q02	Closed constructed response	Connections
M302Q01	Closed constructed response	Reproduction
M438Q01	Closed constructed response	Reproduction
M150Q02	Closed constructed response	Reproduction
M150Q01	Closed constructed response	Reproduction
M155Q01	Open constructive response	Connections
M155Q02	Open constructive response	Connections
M828Q02	Short response	Connections
M155Q04	Complex multiple choice	Connections
M467Q01	Multiple choice	Reproduction
M438Q02	Multiple choice	Reflection
M150Q03	Open constructive response	Connections
M192Q01	Complex multiple choice	Connections
M828Q01	Open constructive response	Connections
M179Q01	Open constructive response	Connections
M828Q03	Short response	Connections
M302Q03	Open constructive response	Reflection
M513Q01	Open constructive response	Connections
M155Q03	Open constructive response	Connections

in standard form. In *connection cluster* of competency, the problem solving is not simply routine; the context is less familiar, and there is the use of interpretation and explaining, handling and relating different systems of representation, selecting and using strategies for solving non standard problems. Finally the *reflection cluster* of competency includes an element of reflectiveness about the processes needed or used to solve a problem. These tasks require understanding, creativity, being able to relate to problem solving and being able to generalize, justify and communicate solutions orally or in writing about mathematical matters (OECD, 2003; Sáenz, 2008). It is not surprising that PISA uses a competency framework that is reminiscent of hierarchical development from simple to advanced forms of cognition since it has some constructivist influences. Conway and Sloane (2006) suggest that situated cognition, Piaget's constructivism and Realistic Mathematics Education (RME) have influenced PISA design of mathematics literacy study. Thus, both the three-tier levels of Friel et al. (2001) and PISA competency levels are constructivism oriented (Conway & Sloane, 2006; Curcio, 1987).

3.1 Two-way taxonomy

It has been shown in Section 2.1, that research suggests a number of possible classifications of graphical comprehension (e.g. Friel et al., 2001; Gal, 1998). To these we can also add PISA's competency classification, as outlined earlier. However, based on the focus of the present study on 'tool use and forms of expression', the students' graphicacy is analyzed using a two-way categorization of interaction with graphical artefacts. This approach is influenced by other studies such as Gal (1998), Roth and Bowen (2001), Ben-Zvi and Arcavi (2001), where there is an indication of a two-way mode of interaction with graphical artefacts. This approach to interaction with graphical artefacts also seems to find traction in Charles S. Peirce's semiotics, where the construct *immediate* and *dynamical* object is espoused in respect to a sign (Peirce,

1958 [CP 8.343]). In the present study, the two-way categorization involves a ‘primordial interaction’; which is basically an identification approach and an ‘advanced interaction’—a critical-analytical approach. In this way, the claim that access to and use of operational tools, domain specific forms of expression and concept are important in graphical literacy, is accounted for. The two categories are: (a) those items considered as requiring an identification approach: these include such items that make use of students’ visual mapping abilities and (b) those items requiring a critical-analytical approach: these include items that require a critical disposition, creative use of mathematical operators and forms of expressions (Fig. 1).

It is noteworthy that following the model outlined by Roth and Bowen (2001), these levels are also not mutually exclusive, as such. The implication of this observation is that in this framework namely the idea of strict hierarchal levels is, to a certain extent, eliminated: the focus is instead on the interaction between the two approaches, thereby increasing the intersection space. By focusing on subject specific tools and forms of expression the framework then points to the need to find suitable teaching and learning processes that promote the interaction between these two approaches. Thus, instead of levels of graphical comprehension, it is feasible from this framework, to talk of the level or the extent of interaction between the identification and critical-analytical approach.

A tabular outline of the application of the framework is provided in Table 2.

In reference to the use of operational tools, domain specific forms of expression and concept, the proposed taxonomy assumes the presence of a dominant tool of operation or concept necessary for effective interaction with a given graphical artefact. Thus, the items were classified based on ‘ease of accessibility’ with respect to subject specific tools and forms of expression (underlying concept) that were perceived as dominating the questions (see Appendix 1 for application of framework).

Some findings from research were employed to determine potential items that might require a critical-analytical approach. For example, research shows that some students, even after several years of schooling, still have problems with ratio and proportions (see e.g., Adjange & Pluvineau, 2007; Noelting, 1980) probability concept, (see e.g., Jones, 2005) aspects of ‘multireferencing’ (Guthrie, et al., 1993), visual-spatial skills (including mental rotation) (see e.g., Stavridou & Kakana, 2008) and the ability to explain their strategies (see e.g., Diezmann & Lowrie, 2009). Aspects of the taxonomy conducted so far were subjected to peer discussions at the Nordic Educational Research Network Trondheim conference in Norway (Olande, 2009). The criterion was also applied to some items from the *Nationella provet*—the Swedish national test in mathematics, where two senior researchers arrived at a general consensus on the classification of items containing graphical elements.

3.2 Aims of the study

The objective of the present study is to establish the general nature of students’ interaction with graphical artefacts (standard of graph comprehension) based on performance patterns

Fig. 1 Categories of graphical interaction

Table 2 Levels of interaction with graphical artefacts

Identification	Critical-analytical
Identifying or reading off values from the graphical system	Critical-analytical interaction with the graphical system: a global rather than a local perspective to data, mental rotation etc.
Manipulation of aspects of the graphical systems using elementary operators e.g. addition and subtraction	Multi-referencing, creative use of mathematical concepts, forms of expression and integrating ideas from other subject domains.
Application of terms and concepts that are non problematic from an everyday perspective.	Appropriate use of argumentation and communication skill to express reflections from the graphical system.

on PISA test items for students from the Nordic countries. In particular the focus will be on the application of subject specific tools or operators and forms of expression. Given the social and cultural influences of use of tools and forms of expression, it is necessary to limit the scope of the study to a cultural homogeneous region (viz., the Nordic countries). It is envisaged that the performance may vary depending on a number of factors such as mathematical context and content (see e.g., Lowrie et al., 2011) as well as the proximity or ease of accessibility of the subject specific tools to students' everyday experience or orientation. The guiding questions for the study are thus:

- What is the pattern of performance for PISA test items containing graphical artefacts viewed from a subject specific tool and form of expression perspective?
- How can students' interaction with items containing graphical artefacts be characterized?

4 Results and analysis

The results are presented in two sections: firstly, the data are presented in relation to the taxonomy outlined in Table 2; secondly, the data are analyzed using cluster analysis. In this way it is possible to compare the distribution of items according to the two methods, these can also be compared with the PISA framework classification as outlined in Table 1.

4.1 Classification of graphical items

Based on the proposed taxonomy provided in Table 2, the selected items containing graphical artefacts are then ordered and classified as shown in Fig. 2.

Generally, based on simple success rate average, Fig. 2 seems to suggest that items considered as requiring identification approach (I) registered higher success rates than the items considered as requiring critical-analytical approach (CA). However, items M155Q04 and M467Q01 seem to register higher success rate scores in relation to other CA items.

4.2 Cluster analysis of graphical items

Having used the proposed taxonomy given in Table 2, further exploration using agglomerative cluster analysis is employed with proximity measurement, determined by *Average linkage*. Cluster analysis is employed here primarily as an independent means of identifying the occurrence of

Fig. 2 Item success rate showing average score for the Nordic countries with a clear distinction for the items e.g. M438Q02 through to M155Q04 are critical-analytical items

natural items grouping from students' performance and secondly as a way of controlling the viability of the proposed two-way taxonomy. Cluster analysis groups individuals or objects in a way that objects in one cluster are more similar to one another than they are to objects in another cluster (Hair, Black, Babin & Anderson, 2010). Since cluster analysis is a multivariate statistical technique, it is worthwhile to state that in the present study, the individual item is perceived as a single observation with the countries score (p -values) being the different dimensions of the observation. Thus, the data can be characterized as involving 19 observations of a 5-dimensional variable i.e. the country item scores for Denmark, Finland, Iceland, Norway and Sweden.

In agglomerative cluster analysis the individual observations or items can be viewed as forming the basic cluster units, this then joins with other items showing similar characteristics until the entire set of items form one major cluster. Thus, it is important to identify relevant clustering levels. Common to most statistical methods, some knowledge of the items or observation used in the study is of vital importance, at least as a way of controlling the reasonability of the clusters formed. In the present study, the determination of relevant cluster level is motivated by the desire for generality and simplicity rather than the desire for details (Romesburg, 1984). Also, 'first-step' cluster levels, though perceived as providing concrete details, may be prone to random fluctuation as evident when different methods are used or when the order of item input changes. For the items used in the present study, two major clusters are identifiable by general inspection. These are clearly determined when the dendrogram is cut at about height 70, giving the cluster formed at the node on approx. heights 48 and approx. height 57 (Fig. 3). These clusters will be referred to here as *cluster I* and *cluster II* respectively. Specifically cluster I includes the items merging at node 48 i.e. starting from the 'top' (far left) of the dendrogram, and is comprised of items M155Q01 moving down to item M302Q02. The remaining items comprised of items M155Q03 down to item M192Q01, form cluster II.

4.2.1 Stability of clusters

Since there is a tendency for the interpretation process of the results from cluster analysis to be subjective, Olsen (2005), suggests a set of criteria focusing on the stability of the clusters

Fig. 3 Dendrogram of the clusters of selected items containing graphical artefacts

formed as a first step in ascertaining validity. Following this suggestion, the clusters formed for the items were subjected to the criteria of (i) internal stability, (ii) stability under selection of technique and (iii) face value. The criterion for external stability was perceived as remotely suggested based on some of the studies suggesting a two-way classification of students' interaction with graphical artefacts (see e.g. Ben-Zvi & Arcavi, 2001; Gal, 1998)—including the application of the categories outlined in Table 2. Internal stability of the clusters was deemed to be fulfilled regarding both external isolation and internal cohesion. The selected technique was also found to provide solutions that were similar to those achieved through *Ward's method* with *Euclidean* measure of proximity. The removal and addition of some of the observations, an exercise suggested by Olsen (2005.), also did not affect the clusters formed: in particular items M150Q02, M155Q02, M155Q03 and M179Q01 were removed since the students' solutions for these items included mixed scores that is, full and partial credit awards (see Appendix 2).

Similar clusters were, to a large extent, also observed at the national level for some of the countries; this observation was taken as implying cluster stability.

Regarding internal cohesion, the results were assessed for uncertainty using AU (approximately unbiased) p -value and BP (bootstrap probability). It is noteworthy that Fig. 4 was produced using *Average linkage* method with *Euclidean proximity measurement*. According to Suzuki and Shimodaira (2006), AU p -values are a better approximation to unbiased p -value than BP values computed by normal bootstrap re-sampling. P -value is a value between 0 and 1, which indicates how strongly the cluster is supported by data: with $nboot$ pegged at 10000 (i.e. the number of bootstraps made), it was observed that the clusters of items formed by the highlighted items in Fig. 4, were strongly supported by the data.

Fig. 4 Cluster dendrogram with AU/BP values showing aspects of internal cohesion of the two clusters of items containing graphical artefacts

For a cluster with AU p -value > 0.95 , the hypothesis that the cluster does not exist is rejected at significance level 0.05. Thus, there is a strong indication that the clusters formed in Fig. 4 are probably authentic; for both cluster I and II, the AU p values are observed to be greater than 0.99. The results from cluster analysis were also controlled for standard errors of the AU p -values by plotting a p -value versus standard error graph—the values were less than 0.008 indicating that the clusters are not significantly influenced by random errors.

5 Discussion

The general aim of the study is to characterize the results of Nordic students' interaction with graphical artefacts based on the PISA survey test (a paper and pen environment). From the two-way taxonomy of interaction with graphical artefacts (see Fig. 1), it is observed that the general pattern of performance is similar for all the Nordic countries. Students registered high success rates for items considered to require an identification approach while the same is not entirely true for items perceived to require a critical-analytical approach. Whereas this finding may seem rather obvious given that challenging questions are expected to record low success rates, it is observed that this suggestion might not necessarily apply to the three-tier competency classification given by PISA of increasing levels of item complexity from reproduction-connection-reflection. A number of items from the connection cluster are observed in both identification as well as critical-analytical item categories (see Fig. 3 and Table 1). In this regard the proposed taxonomy might be perceived as bringing clarity to the PISA classification (with regard to category or data reduction). Indeed, it is observed that items from the PISA's reflection cluster fall under the critical-analytical category of items, while items from the reproduction clusters of items fall under the identification category. The utility of cluster analysis is also observed when considering that while Åberg-Bengtsson (2005) was able to identify a subset of SweSAT that measured analytical dimension (dealing with calculation of sort), the present study shows the existence of differential performance based on the kind of calculation involved with the tasks. The construct involving a two-way interaction with graphical artefacts that is, an identification and critical-analytical approach, is perceived as viable and has, to a large extent, been confirmed through cluster analysis. The results with cluster analysis show that the selected items also coalesces into two major clusters with most items requiring an identification approach coalescing in what is in the present study referred to as cluster I.

We have indicated earlier that 'item complexity' is a function of a number of factors such as 'proximity' of the students to the tools and forms of expression demanded by the task and the choices they make in tool selection in engaging a graphical artefact. From the dendrogram provided in Fig. 3, it is observed that some of the items classified as requiring a critical-analytical approach are clustered with those deemed as requiring an identification approach. This provides an opportunity to look closer at the items in light of the classification criteria. Item M155Q04 (population pyramids: it is not yet in the public domain) was classified as CA on account of its demanding the consultation of several graphical artefacts, however most of the question tasks were such that they required the identification of values and trends. The question was also of a multiple choice format. Item M467Q01 was classified as CA since it required calculation with proportions; this was also a task of a multiple choice format. A possible explanation for this discrepancy is that the item format might have reduced the burden on some of the choices the students had to make, that is, they might have had a scaffolding effect in terms of application of subject specific tools and forms of expression which contributed to the higher performance. The 'difference' between results from cluster analysis and the proposed construct can also be seen as revealing the dynamics or the

interplay of the students' approach to interaction with graphical artefacts. As outlined earlier the two approaches to graphicacy, viz.: identification and critical-analytical, ought not to be seen as mutually exclusive. Whereas the proposed construct is 'static'. Through cluster analysis it may be possible to determine which subject specific tools and operators have been integrated into the students' repertoire for engaging graphical artefacts, for example, in this case one might need to re-examine the justification of considering proportional reasoning as a problem area for students with several years of schooling. This observation can also be perceived as an advantage of using cluster analysis in combination with the proposed taxonomy.

From the classification of students' results for the Nordic countries, it is observed that performance on items considered as requiring critical-analytical approach was lower than those of items considered as requiring identification approach. This observation could be an indicator that students seem to be more confident using aspects involving visual dimensions or simple procedural operations. Given that graphical artefacts can easily be used to present facts in distorted ways for egoistic ends, it is important to develop a critical-analytical approach to data. Students need to be nurtured into ways that interrogate the data both critically and analytically—this becomes more urgent since the methods of presenting distorted information in graphical representative are becoming more sophisticated thanks to the burgeoning information age. In this regard the names given to the proposed constructs also seem to effectively communicate the required prerequisite for effective citizenry.

6 Conclusion

This study of students from the Nordic countries sought to characterize their performance on items containing graphical artefacts such as bar graphs, pie charts and line graphs. The general observation from both the analytical construct of the study and the results from cluster analysis showed that performance on items containing graphical artefacts crystallized into two groups; one group consisting of items demanding identification skills and the other group consisting of items demanding critical-analytical skills. While this observation appears to lend credence to the thesis which supports the presence of two modes of interactions with graphical artefacts, it is noteworthy that the present study is limited to the Nordic countries. Thus, there is need for similar research from other regions in search for a global pattern of interaction with graphical artefacts. Significantly, the focus on the use of subject specific tools or operators and forms of expression recognizes the importance of being able to identify and select appropriate 'external' resources for effective interaction with graphical artefacts and thus underscores the need for teaching and learning activities to develop and nurture students' critical-analytical disposition. It is essential that critical-analytical competences are trained to give students a chance to develop a deeper understanding of graphical artefacts and a critical attitude to claims presented in sometimes complex or even misused diagrams. The taxonomy also highlights the role of tools and forms of expression calling on teachers' to be vigilant on the same in the design and use of tasks (e.g., assessment tasks) containing graphical artefact in teaching and learning situations.

Appendix 1

Below are some of the items used in the study and which have been released into the public domain by PISA. We also illustrate how the proposed taxonomy has been applied. The bottom line is identifying items that require the application of simple visual inspection, elementary tools and forms of expression.

M150Q01-3**YOUTH GROWS TALLER**

In 1998 the average height of both young males and young females in the Netherlands is represented in this graph.

Question 1: GROWING UP

M150Q01- 0 1 9

Since 1980 the average height of 20-year-old females has increased by 2.3 cm, to 170.6 cm. What was the average height of a 20-year-old female in 1980?

Answer:cm

Question 2: GROWING UP

M150Q02- 00 11 21 22 99

According to this graph, on average, during which period in their life are females taller than males of the same age?

.....

Question 3: GROWING UP

M150Q03- 01 02 11 12 13 99

Explain how the graph shows that on average the growth rate for girls slows down after 12 years of age.

.....

- **M150Q01**: this item is coded as identification (I) since the solution can be derived by visual inspection, picking the solution from the sentence and also by reading the values.
- **M150Q02**: this item is coded as identification (I) since the solution can be derived by visual inspection, that is to say, by pointing out values from the graphical artefact.
- **M150Q03**: whereas aspects of this item are accessed visually, students seem to have difficulties explaining or articulating their solution strategies, thus it is classified as critical-analytical (CA).

M179Q01

- **M179Q01**: this item is classified as CA since it requires insight into the fact that it might be difficult to make adequate comparisons with limited data; the student is called upon to realize the importance of aspects not presented in the graphical representation, it also requires the justification of solution strategies; an observation that might prove to be challenging to most students.

M438Q01-2**EXPORTS**

The graphics below show information about exports from Zedland, a country that uses zeds as its currency.

Total annual exports from Zedland in millions of zeds, 1996-2000

Distribution of exports from Zedland in 2000

Question 1: EXPORTS

M438Q01 - 0 1 9

What was the total value (in millions of zeds) of exports from Zedland in 1998?

Answer:

Question2: EXPORTS

M438Q02

What was the value of fruit juice exported from Zedland in 2000?

- A 1.8 million zeds.
- B 2.3 million zeds.
- C 2.4 million zeds.
- D 3.4 million zeds.
- E 3.8 million zeds.

- **M438Q01**: this item is coded as I since the solution can be derived by visual inspection; picking the solution from the bar graphs.
- **M438Q02**: this item is coded as CA since the solution requires knowledge of proportions and calculation with percentage.

M467Q01**Question 1: COLOURED CANDIES**

M467Q01

Robert's mother let him pick one candy from a bag. He can't see the candies. The number of candies of each colour in the bag is shown in the following graph.

What is the probability that Robert will pick a red candy?

- A 10%
- B 20%
- C 25%
- D 50%

- **M467Q01**: this item is coded as CA; it is a task with multiple solution strategies however, the general demand requires knowledge or appreciation of proportional calculation and percentages [and probably some elementary knowledge of the concept of probability]

M513Q01**Question 1: TEST SCORES**

M513Q01 - 0 1 9

The diagram below shows the results on a Science test for two groups labelled as Group A and Group B.

The mean score for Group A is 62.0 and the mean for Group B is 64.5. Students pass this test when their score is 50 or above

Looking at the diagram, the teacher claims that Group B did better than Group A in this test.

The students in Group A don't agree with their teacher. They try to convince the teacher that Group B may not necessarily have done better.

Give one mathematical argument, using the graph, that the students in Group A could use.

- **M513Q01:** this item is classified as CA since it requires insight into the statistical concept of variability [a concept that, according to research, is challenging to most students]. The task also requires the justification of solution strategies; an undertaking that might prove challenging to most students.

Appendix 2

Here a dendrogram shows selected items that received full credit: In this case the items that have been removed are M150Q02, M155Q02, M155Q03 and M179Q01 (Fig. 5).

Fig. 5 Dendrogram of the clusters exclusively awarded full points. Thus, pointing to the stability of the clusters formed

References

- Åberg-Bengtsson, L. (2005). Separating the quantitative and analytic dimensions of the Swedish Scholastic Aptitude Test. *Scandinavian Journal of Educational Research*, 49, 359–383.
- Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching*, 43, 43–62.
- Adjiage, R., & Pluvineau, F. (2007). An experiment in teaching ratio and proportion. *Educational Studies in Mathematics*, 65, 149–175.
- Balchin, W.G.V., & Coleman, A.M. (1965). *Graphicacy should be the fourth pack in the ace*. Times educational supplement November, 5.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 36–65.

- Bergqvist, T., Lithner, J., Sumpter, L. (2003). Reasoning characteristics in upper secondary school students' task solving. *Research reports in mathematics education, 1*, Umeå: Mathematics Department, Umeå University.
- Bertin, J. (1983). *Semiology of graphics: Diagrams networks maps*. (W. J. Berg, trans.) Madison: The University of Wisconsin Press. (Original work published 1967).
- Conway, P., & Sloane, F. (2006). *International trends in post-primary mathematics education: Perspectives on learning, teaching and assessment*. Dublin: NCCA.
- Costiagn-Eave, P., & Macdonald-Ross, M. (1990). William Playfair (1759–1823). *Statistic Science, 5*, 318–326.
- Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382–393.
- Daniels, H. (2008). *Vygotsky and research*. London: Routledge.
- Diezmann, C., & Lowrie, T. (2009). Primary students' spatial visualization and spatial orientation: An evidence base for instruction. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 417–424). Thessaloniki: PME.
- Dudaite, J., & Eljio, A. (2008). *Trends in similarities and differences of students' mathematics profile in various countries*. Paper presented in the proceedings of the 3rd IEA International Research Conference. Taipei, China Retrieved June 15, 2012 from http://www.iea.nl/fileadmin/user_upload/IRC/IRC_2008/Papers/IRC2008_Dudaite_Eljio.pdf.
- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124–158.
- Funkhouser, H. G. (1937). Historical development of the graphical representation of statistical data. *Osiris, 3*, 269–404.
- Gal, I. (1998). Assessing statistical knowledge as it relates to students' interpretation of data. In S. P. Lajoie (Ed.), *Reflections on statistics: Learning, teaching, and assessment in grades K-12* (pp. 275–295). Mahwah NJ: Lawrence Erlbaum Associates.
- Goodchild, S., & Grevholm, B. (2007). An exploratory study of mathematics test results: What is the gender effect? *International Journal of Science and Mathematics Education, 7*, 161–182.
- Gustafsson, J.-E., Wedman, I., & Westerlund, A. (1992). The dimensionality of the Swedish Scholarstic Aptitude Test. *Scandinavian Journal of Educational Research, 36*, 21–39.
- Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive process and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology, 18*, 186–221.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. Upper Saddle River: Pearson Prentice Hall.
- Jones, G. W. (2005). *Exploring probability in schools: The challenges for teaching and learning*. New York: Springer.
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding groups in data: An introduction to cluster analysis*. Hoboken: John Wiley and Sons.
- Kjærnsli, M., & Lie, S. (2004). PISA and scientific literacy: Similarities and differences between the Nordic countries. *Scandinavian Journal of Educational Research, 48*, 271–286.
- Lie, S., & Roe, A. (2003). Exploring unity and diversity of Nordic reading literacy profiles. In S. Lie, P. Linnakylä, & A. Roe (Eds.), *Northern lights on PISA. Unity and diversity in the Nordic countries in PISA 2000* (pp. 147–157). Oslo: Department of Teacher Education and School Development, University of Oslo.
- Lowrie, T., & Diezmann, C. M. (2007). Solving graphics problems: Students performance in the junior grades. *The Journal of Educational Research, 100*, 369–377.
- Lowrie, T., Diezmann, C. M., Logan, T. (2011). Understanding graphicacy: Students' making sense of graphics in mathematics assessment tasks. *International Journal for Mathematics Teaching and Learning*. Retrieved January 12, 2012 from <http://www.cimt.plymouth.ac.uk/journal/default.htm>.
- Noelting, G. (1980). The development of proportional reasoning and the ratio concept. Part 1—Differentiation of stages. *Educational Studies in Mathematics, 11*, 217–253.
- OECD. (2003). *PISA 2003 assessment framework—mathematics, reading, science and problem solving knowledge and skills*. Paris: Author.
- OECD. (2007). *PISA 2006: Science competencies for tomorrow's world executive summary*. Paris: Author.
- Olande, O. (2009). *Analysis of students' performance in graphicacy from PISA test items*. Norway: Nordic Educational Research Network Trondheim.
- Olsen, R.V. (2005). *Achievement tests from an item perspective. An exploration of single item data from the PISA and TIMSS studies, and how such data can inform us about students' knowledge and thinking in science*. PhD-thesis, University of Oslo. Oslo: Unipub
- Olsen, R. V. (2006). A Nordic profile of mathematics achievement: Myth or reality? In J. Mejding & A. Roe (Eds.), *Northern Lights on PISA 2003—a reflection from the Nordic countries* (pp. 33–45). Copenhagen: Nordic Council of Ministers.

- Olsen, R. V., & Grønmo, L. S. (2006). What are the characteristics of the Nordic profile in mathematical literacy? In J. Mejdning & A. Roe (Eds.), *Northern lights on PISA 2003—a reflection from the Nordic countries* (pp. 47–57). Copenhagen: Nordic Council of Ministers.
- Österholm, M. (2008). Do students need to learn how to use their mathematics textbooks? The case of reading comprehension. *Nordic Studies in Mathematics Education*, 13, 53–73.
- Peirce, C. S. (1958). In A. W. Burks (Ed.), *Collected papers of Charles Sanders Peirce* (Vol. VII–VIII). Cambridge, MA: Harvard University Press. Quoted by volume followed by paragraph number.
- Postigo, Y., & Pozo, J. I. (2004). On the road to graphicacy: The learning of graphical representation systems. *Educational Psychology*, 24, 623–644.
- Radford, L. (2003). On culture and mind. A post-Vygotskian semiotic perspective, with an example from Greek mathematical thought. In M. Anderson, A. Sáenz-Ludlow, S. Zellweger, & V. Cifarelli (Eds.), *Educational perspectives on mathematics as semiosis: From thinking to interpreting to knowing* (pp. 49–79). Ottawa: Legas Publishing.
- Radford, L. (2008). The ethics of being and knowing. In L. Radford, G. Schubring, & F. Seeger (Eds.), *Semiotics in mathematics education* (pp. 215–234). Rotterdam: Sense.
- Radford, L. (2009). Signifying relative motion: Time, space and the semiotics of cartesian graphs. In W.-M. Roth (Ed.), *Mathematical representation at the interface of body and culture* (pp. 45–69). Charlotte, NC: Information Age Publishing.
- Raschke, R. L., & Steinbart, P. J. (2008). Mitigating the effects of misleading graphs on decisions by educating users about the principles of graph design. *Journal of Information Systems*, 22, 23–52.
- Robinson, A. H. (1967). The thematic maps of Charles Joseph Minard. *Imago Mundi*, 21, 95–108.
- Roe, A., & Taube, K. (2006). How can reading abilities explain differences in maths performances? In J. Mejdning & A. Roe (Eds.), *Northern lights on PISA 2003*. Copenhagen: Nordic Council of Ministers.
- Romesburg, H. C. (1984). *Cluster analysis for researchers*. Belmont, CA: Lifetime learning publications.
- Roth, W.-M. (2003). *Towards an anthropology of graphing*. The Netherlands: Springer.
- Roth, W.-M. (2008). The dawning of signs in graph interpretation. In L. Radford, G. Schubring, & F. Seeger (Eds.), *Semiotics in mathematics education* (pp. 83–102). Rotterdam: Sense.
- Roth, W.-M. (2009). Embodied mathematical communication and the visibility of graphical features. In W.-M. Roth (Ed.), *Mathematical representation at the interface of body and culture* (pp. 95–121). Charlotte, NC: Information Age Publishing.
- Roth, W.-M., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in Mathematics Education*, 32, 159–194.
- Sáenz, C. (2008). The role of contextual, conceptual and procedural knowledge in activating mathematical competencies (PISA). *Educational Studies in Mathematics*, 71, 123–143.
- Sfard, A. (2001). There is more to discourse than meets the ears: Learning from mathematical communication things that we have not known before. *Educational Studies in Mathematics*, 46, 13–57.
- Stavridou, F., & Kakana, D. (2008). Graphic abilities in relation to mathematical and scientific ability in adolescents. *Educational Research*, 50, 75–93.
- Suzuki, R., & Shimodaira, H. (2006). *Pvclust*. Retrived January 16, 2012 from <http://www.is.titech.ac.jp/~shimo/prog/pvclust/>.
- van Oers, B. (2000). The appropriation of mathematical symbols: A psychosomatic approach to mathematical learning. In P. Cobb, E. Yackel, & K. McClain (Eds.), *Symbolizing and communicating in mathematics classrooms: Perspectives on discourse, tools, and instructional design* (pp. 133–176). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Vygotsky, L. S. (1978). In M. Cole, V. John-Steiner, S. Scribner, & Souberman (Eds.), *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, 21, 14–23.
- Wainer, H. (2009). *Picturing the uncertain world: How to understand, communicate, and control uncertainty through graphical display*. New Jersey: Princeton University Press.
- Wertsch, J. V. (1985). *Vygotsky and the social formation of the mind*. Cambridge, MA: Harvard University Press.
- Winn, W. (1993). An account of how readers search for information in diagrams. *Contemporary Educational Psychology*, 18, 162–185.