<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper published in *Studies in Higher Education*. This paper has been peer-reviewed but does not include the final publisher proof-corrections or journal pagination.

Citation for the original published paper (version of record):

Forsman, J., Linder, C., Moll, R., Fraser, D., Andersson, S. (2014)

A new approach to modelling student retention through an application of complexity thinking.

Studies in Higher Education, 39(1): 68-86

<http://dx.doi.org/10.1080/03075079.2011.643298>

Access to the published version may require subscription.

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-170290>

This paper has been published by Studies in Higher Education as: Forsman, J., Linder, C., Moll, R., Fraser, D. and Andersson, S. (2014). A new approach to modelling student retention through an application of complexity thinking. Vol. 39, no 1, 68-86. The published copy can be obtained from <http://dx.doi.org/10.1080/03075079.2011.643298>

Published online: 02 Feb 2012

A New Approach to Modelling Student Retention Through an Application of Complexity Thinking

Jonas Forsman^a, Cedric Linder^{ab}, Rachel Moll^c, Duncan Fraser^d, and Staffan Andersson^a

^a Division of Physics Education Research, Department of Physics and Astronomy, Uppsala University, Sweden

^b Department of Physics, University of the Western Cape, South Africa

^c Faculty of Education, Vancouver Island University, Canada.

^d Department of Chemical Engineering, University of Cape Town, South Africa

Abstract: Complexity thinking is relatively new to education research and has rarely been used to examine complex issues in physics and engineering education. Issues in higher education such as student retention have been approached from a multiplicity of perspectives and are recognized as complex. The complex system of student retention modelling in higher education was examined to provide an illustrative account of the application of complexity thinking in educational research. Exemplar data was collected from undergraduate physics and related engineering students studying at a traditional Swedish university. The analysis shows how complexity thinking may open up new ways of viewing and analysing complex educational issues in higher education in terms of nested, interdependent and interconnected systems. Whilst not intended to present new findings, the paper does illustrate a possible representation of the system of items related to student retention and how to identify such influential items.

Keywords: student retention; modelling systems; complexity thinking

Introduction and Research Aim

Complexity thinking, which is derived from complexity theory, is a powerful conceptual framework in education that draws on the qualities of complex systems to characterize learning systems (e.g. Davis and Sumara 2006). As such, with its organic, non-linear, relational and holistic features, complexity thinking presents a stark point of departure for contemporary educational research thinking (Morrison 2006). Thus, its application in education research is still relatively rare, particularly in higher education contexts, and a variety of good exemplars are hard to find for newcomers contemplating the use of complexity thinking as a conceptual framework. This article uses the field of student retention to provide such an exemplar.

When using complexity thinking, qualities such as decentralized network structure and short-range communication between agents can be seen to facilitate emergence of order in seemingly chaotic complex systems. Student retention is a complex system

comprised of variables such as curriculum, financial and social supports where learning entails developing a sense of belonging and a desire to continue and complete a program (cf. Bean 2005; Tinto 2010; and see examples in Braxton 2000). Thus, we would argue that complexity thinking, as a trans-disciplinary perspective, is ideal for student retention modelling. Complexity thinking explicitly recognizes the connectedness and interactions between the multitude of variables that are currently recognized as influential to student retention (for example, see Bean 1982; and Tinto 1997). Our aim in this article is to illustratively explore the potential advantages of applying complexity thinking by bringing in exploratory factor analysis and multidimensional scaling to explore the problematic issue of student retention. Specifically we hope to illustrate how complexity thinking may be able to identify emerging structures and dynamics such as the adaptive and decentralized system of variables that influence student retention in higher education.

To achieve this aim we have used exemplar data consisting of retention questionnaire responses completed by Swedish first year physics and related engineering students. This particular context was chosen because currently engineers and scientists are in demand (for example, see European Commission 2004; Committee on Science, Engineering, and Public Policy 2007), while poor retention rates are also an international issue (see, for example Statistics Sweden and National Agency for Higher Education 2003, 2005, 2007, 2009, 2010). This is a complex context for applying student retention models that have been developed in the literature.

The data set was not selected in a way aimed at producing new variables needed for modelling student retention, but to provide a theoretical and illustrative account of the application of complexity thinking.

Historical overview of models of student retention

We have made a case for using an illustrative student retention data set, and to proceed a summative historical overview of models of student retention is needed. Readers who want a more extensive overview are directed to Summerskill (1962), Tinto (1975), Braxton (2000) and Metz (2004).

Throughout the history of building insightful models of student retention, the notion of *complexity*¹ has been apparent, but it has not been brought to the fore in previous model designs. For example, Spady (1971, 38) argues that the formulation of a truly comprehensive model of student retention needs a perspective that ‘regards the decision to leave a particular social system [studies in higher education] as the result of a complex social process’. More recently Bean (2005, 238) has argued that ‘students’ experiences are complex, and their reasons for departure are complex’. There are many other examples, see Spady (1970, 1971), Cabrera, Nora, and Castañeda (1993), Tinto (2010), Yorke and Longden (2004), and Barnett (2007).

This brings us to our starting point for the theoretical development for this study. Currently there are two models of student retention that are widely used: The Student Integration Model (Tinto 1975, 1987, 1997) and The Student Attrition Model (Bean

¹ Complexity should not be seen as a synonym for ‘complicated’, but rather as a kind of ‘generative metaphor’ (Schön 1983) that extends a characterization of ‘a complex unity that is capable of adapting itself to the sorts of new and diverse circumstances that an active agent is likely to encounter in a dynamic world’ (Davis and Sumara 2006, 14).

1980, 1982). While these models are seen by some researchers as being two separate, and even independent systems, we concur with Cabrera et al. (1992, 145) that both 'regard persistence [retention] as the result of a complex set of interactions over time.'

Yorke and Longden (2004) describe how the focus of early studies of student retention in higher education were on university structures, for example, libraries, schedules, courses, or examination timetables. Thereafter began a shift in modelling student retention towards incorporating a social integration perspective, influenced largely by the work of Spady (e.g. Spady 1970, 1971).

According to the social integration perspective, becoming integrated within a social system requires learning the norms, value-systems, and beliefs through interactions within the system. The social integration perspective played a major role in the development of Spady's theoretical model (Spady 1970, 1971); students needed to become a part of the social world of the university if the departure rates were to decline. In this model, social integration is a process that encompasses many aspects of students' everyday lives. This includes friendships, family support, the students' feeling of satisfaction and intellectual development, and so forth. Spady's model also included student characteristics such as grade performance, family background, and academic potential.

The social integration perspective gained momentum in student retention research through its potential for informing students' and universities' actions toward working to retain more students. Tinto (1975) published an expanded version of Spady's model. Tinto made a distinction between the social system of the university and the academic system, and argued that students also need to become academically integrated to persist in their studies. He posited that some interactions that lead to social integration, for example, making friends with fellow students, do not necessarily lead towards integration into the academic system of the university. The academic system, according to Tinto's (1975) conceptual framework, contains the academic rules, norms and expectations that govern academic interaction within the given institutional context.

During the early 1980s, researchers in the field empirically tested Tinto's constructs, and found that many of them were indeed impacting student retention. At this time, Bean, drawing on a psychological background, critiqued Tinto's model for its lack of external factors – such as economy and housing (Bean 1980). The point of departure for Bean's (1980) model was that student attrition should be seen as analogous to work turn-over in a traditional employment setting. Factors such as social experiences, the experience of the quality of the university, and family approval shape students' attitudes and behaviour within the university context.

To evaluate Bean's and Tinto's models, Cabrera et al. (1992) surveyed 2453 full-time American freshman students. Their findings indicate that the two student retention models have common ground and that they support each other in explanatory value. The questionnaire they designed was made up of 79 items, selected from well-validated instruments previously used in the field of student retention (for example, see Bean 1982, Pascarella and Terenzini 1979).

Later, Eaton and Bean (1995) theorized that students' experiences shape their individual behavioural approaches towards university life. This development expanded their earlier model of student attrition by adding approach and avoidance behavioural theory. Thus some students' experiences lead towards avoidance behaviour, and some

towards an approach behaviour, both of which affect academic integration and thus the students' intention to leave or stay.

Tinto (1997) then undertook a case study that led him to expand his model by introducing the notion of 'internal' and 'external' communities that affect student integration into the social and academic systems of the university. He asserted that within classrooms there are 'internal' learning communities where both social and academic systems coexist. Through the concept of learning communities, together with the presence of 'external' communities, new constructs became available that could empower teachers who wanted to improve student retention (Tinto 1997).

After the development of Bean's and Tinto's models, Braxton's contribution to the field has led to a wide recognition of the range of variations within the empirical findings associated with Tinto's model and that it should be 'seriously revised' (Braxton 2000, 258). Braxton and Lien (2000) compiled empirical results on academic integration and concluded that Tinto's claim (1975, 1997) that it is a central construct has yet to be demonstrated empirically. The social integration piece of Tinto's theory has also been examined. Braxton and Hirshy (2004) provided empirical data to support their proposal to incorporate three additional factors that may influence social integration: commitment of the institution to student welfare, institutional integrity, and communal potential. Braxton (2000) suggested that a new foundation for such a modelling system needs to be developed and Tinto (2010) himself argued for the need to develop models that aim towards informing the institutional action of universities.

The next step is then to put forward a modelling system that can adapt to variations within empirical findings, has the ability to harbour constructs of the earlier models, and empowers universities in their actions toward enhancing student retention. The conceptual framework of this article, complexity thinking, is a perspective that can help achieve these aims and has the potential to suggest changes to educational practice.

Conceptual framework

In this section, we will present the concepts that we draw upon from complexity thinking that illustrate the connected, adaptive, and dynamic nature of the complex system of student retention in higher education. At the same time this illustration can be seen to exemplify how drawing on complexity thinking can allow for the generation of a more powerful and holistic modelling system of student retention. To do this we use exploratory factor analysis, multidimensional scaling, and network theory.

Complexity thinking

Complexity thinking aims to describe and understand complex systems and their capacity to show order, patterns, and structure. Especially important is how these orders, patterns and structures seem to emerge spontaneously from interactions between components of systems. Complexity thinking has emerged and taken root in a wide range of disciplines, generating a theory that essentially 'transcends disciplines' (Waldrop 1992). For more details on the historical development of complexity thinking, see Waldrop (1992), and for an overview of current applications of complexity thinking in a wide array of fields, see Mitchell (2009).

Complexity thinking is often pitted against 'classical science', which is, in turn, portrayed in terms of efforts to condense phenomena into their simplest components.

However, to obtain a reasonable portrayal of a complex phenomenon, an understanding of the properties of the components alone is not sufficient. Thus, what is central in describing or understanding a complex system is identifying the components, their interactions, and what *emerges* from the complex system: system behaviours, properties and structures, or the ‘structuring structures’ (Bourdieu 1984) of the complex system (for example, see Davis and Sumara 2006).

One can conceptualize the essential aspects of a complex system’s structure, dynamics, and predictability through metaphors (for example, see Gilstrap 2005), computer simulations (for example, see Brown and Eisenhardt 1997) and systems of modelling (for example, see Mowat and Davis 2010). From this perspective, the essential aspect of complex systems, and what has given rise to complexity thinking’s ubiquitous emergence, is that all complex systems share similar structure and dynamics. The behaviour of complex systems, such as society, organisms, or the Internet can only be conceptually discussed as somewhere in-between complete order and complete disorder. Any attempt to measure or distinguish one system as more complex than another often breaks down (Mitchell 2009). If a system is to be identified as being a complex system what needs to be investigated are the characteristic structures and dynamics that are common among complex systems, not the complexity itself (Davis and Sumara 2006).

The structure of complex systems

Complex systems have *decentralized networked structure*, which means that there are a few components or nodes that are much more connected than others. This kind of structure can be contrast to two other types of networks: (1) centralized networks with only one central node that every other node is connected to; and, (2) distributed networks where all nodes have the same connectivity in the network. Information is spread efficiently in centralized networks, but they are vulnerable to break down due to the dependency on the central node. On the other hand, distributed networks are robust but inefficient in spreading information. In decentralized networks when a highly connected component is removed or breaks down, the whole system will suffer considerable damage. A decentralized system will remain stable, however, with the removal of any of the many less important or less connected nodes.

Due to their decentralized structure, all complex systems are *networked* with other complex systems. Moreover, components within a complex system can be considered to be complex systems themselves, thus complex systems are *nested*. Nested systems have similar structure and dynamics but operate on different scales (time, size and so forth). For example, mathematics learning-for-teaching has been modelled as several nested systems: subjective understanding, classroom collectivity, curriculum structure, and mathematical objects (Davis and Simmt 2006). Each level of such nested complex systems exhibits similar structures and dynamics, but operates within different time-scales (for example, subjective understanding has a faster rate of change than mathematical objects) and represents a different level of analysis (such as the level of an individual, the level of a group of individuals, or the level of a particular culture).

Dynamics of complex systems

One key aspect of complex systems is that they are continually changing as the components in the system interact with the external environment and with one another.

This means that complex systems are *adaptive* and *self-organizing*; properties, behaviour and structure all emerge without an external system or an internal ‘leader system’ that controls the complex system.

Components of complex systems interact mainly locally via *neighbour-interactions*, which can fuel processes that lead to emergence such as positive feedback (brings the system to a non-equilibrium state) or negative feedback to maintain equilibrium. Positive feedback tends to amplify, and negative feedback tends to dampen properties, behaviours and structures. Depending on how ‘connected’ each component is with other components within the system, the positive feedback can give rise to the possibility of *emergence*. Complexity thinking has established that decentralized network structure is a key element in facilitating emergence in complex systems. Through the concept of neighbour-interactions and the decentralized network structure we can argue that nested systems that are highly connected can be seen as close to each other (Davis and Sumara 2006).

Complexity thinking is not characterized by a particular method, but by a methodological perspective that employs a range of methods to study complex phenomena (Davis and Sumara 2006). Complex systems are networked constellations of components, which in our example are the students’ viewpoints of their experience of higher education in the first year. Each item in our analysis, such as students’ attitudes towards their program and their financial stability, is considered to emerge from and be situated within multiple complex systems. From here a formulation of a possible visualization of the structure and dynamics of the complex system of student’ retention is created through the use of the following tools that can be seen to be complimentary to those used in complexity studies: exploratory factor analysis, multidimensional scaling, and network theory.

Exploratory factor analysis

Exploratory factor analysis is used to study patterns and order within complex data by comparing angles between points in a multidimensional space. A useful way to view exploratory factor analysis is to see it as essentially what Hoftstede et al. (1990, 299) has called ‘ecological factor analysis’; an analysis where the stability of the analysis does ‘...not depend on the number of aggregate cases but on the number of independent individuals who contributed to each case’.

The items used in the analysis are the retention questionnaire responses plus other student-specific information. Exploratory factor analysis identifies those items that have ‘commonalities’ (Kim and Mueller 1978) by using the covariance between the items. Items with higher covariance are grouped into a number of factors, with the number being determined by the groupings that arise. Using a complexity thinking perspective, these factors were interpreted as a self-organized pattern of nested systems that make up the complex system of student retention in higher education.

Exploratory factor analysis will normally reveal that some of the items are present in more than one of the factors. This was interpreted as evidence of interactions between the nested systems through their shared items.

Multidimensional scaling

As denoted by the conceptual framework, components of a complex system interact locally (Davis and Sumara 2006) and thus components that have a high relative closeness

to other components in the multidimensional scaling analysis can be regarded as being *connected*² and within each other's *zone of influence*. In the multidimensional scaling analysis of the questionnaire data, the answers and their proximities are used to create a representation of the emergent network structure of the complex system. The items may be seen as vertices connected by edges, which form a basis for visualization and allow for measurements of item interaction through the use of network theory.

A good way to determine the relative proximity of items to one another is to use multidimensional scaling because it offers a way to calculate the distances between points of data in multidimensional space. The relative *closeness* ('multidimensional proximity') of items to one another is the 'likeness' or 'similarity' (Schiffman et al. 1981) of those items.

Network theory

The orienting emphasis in network theory is 'structural relations' (Knocke and Yang 2008). From such a framing the essential elements of a network are the nodes (vertices) and the links (edges) between nodes. In the current study, nodes are items examined in the retention questionnaire and student-record data. Network theory is thus a powerful analytic tool to explore and illustrate the connectivity of the structure produced by the multidimensional scaling.

Network theory concepts

The *nodes* represent the components of a network (i.e. items on the retention questionnaire and student-record data), and the edges represent the relationships between the nodes. When two nodes are directly connected the two nodes are said to be adjacent. A *path* is a way through a sequence of nodes that begins with a starting node, follows adjacent nodes through the network and ends at an ending node. When every node in the network is reachable (i.e. a path exists between every node) the network is connected. If there are many paths between two nodes, the *shortest path* between them is the one with the fewest connections made through other nodes (Freeman 1978). Visualization and analysis of networks, and therefore complex systems, is made possible by using these constructs of network theory.

Network measurements and interpretation

The network to represent the system of student retention was formed using multidimensional scaling. As in the literature, we assumed that we had an undirected network where the connections between the nodes did not have a specific direction of influence. Analysis of the created network was done by using Statnet (Handcock et al. 2003), a free package designed for analysing networks which employs the "R" statistical computing and graphics program. Identification of 'important' nodes was done by analysing each node's *betweenness centrality* and *closeness centrality*.

In this study we distinguish between *closeness centrality* and *betweenness centrality* (Bernhardsson 2009). Closeness centrality is an ordinal measure of how 'close' every

² 'Connected' is used as a broad term that encompasses the interaction, communication, and dependence between the different components of the system.

other node is, and it is calculated through finding the shortest path between nodes. Information can be spread to the whole network more efficiently from nodes with high closeness centrality (Freeman 1978). Betweenness centrality is the frequency that one particular node is a part of the shortest path between every other node. Nodes that are more frequently a part of the shortest path between nodes may be interpreted as having a high degree of '*control* of communication' (Freeman 1978, 224) in the network.

Method

The exemplar data was collected from two sources: student-records were used to obtain student demographic information items such as Age, Gender, Retention, and Higher Education Credits achieved within (HECwP) and outside (HECoP) the programme. Thereafter, a questionnaire was developed to explore influences on student retention.

Retention questionnaire

To generate data that we felt would well serve our research aim we chose the following setting: a typical first year physics course at a highly regarded traditional Swedish university (most of the students enrolled were in an engineering programme). For our data collection a questionnaire was constructed, based on previous student retention research and physics education research. In particular the questions that provide the highest explanatory value in Cabrera et al.'s (1992) survey of the convergence of the Student Integration Model (Tinto 1975, 1987, 1997) and the Student Attrition Model (Bean 1980, 1982) were used and some questions were added. See Appendix 1 for the detailed source(s) for each question, plus a detailed motivation for questions that were changed or added. Students answering the questionnaire were asked to mark their level of agreement (or disagreement) with 29 statements on a five-point Likert scale.

To get a good questionnaire-completion response rate we chose a venue that allowed us to easily give a motivating and thorough discussion of our aims and the associated ethical considerations we would uphold. Most (51) students agreed to participate and the questionnaire was administered towards the end of their first year (second semester) of university study in 2009. Thirty-two of the participating students were registered in a four and a half year Master of Science in Engineering (Physics) programme, twelve were registered in a three year Bachelor of Science (Physics) programme and the remaining seven were registered in a four and a half year Master of Science in Engineering (Materials Physics) programme. Re-enrolment in the second year (third semester) was used as a measurement of student retention and was found to be 82.4%.

Illustrative results

Exploratory factor analysis

Having satisfied ourselves that we had data items grounded in the literature, we started with an exploratory factor analysis. This was to identify the nested complex systems that make up the greater system of student retention through the identification of the factors of the overall system. Our analytical tool was the Statistical Package for the Social Sciences, SPSS (Predictive Analytic SoftWare, PASW, version 18.0). Our starting point was the normalized matrix of the questionnaire data together with the students' higher

education credits achieved within and outside their programme, retention (re-enrolment in the second year), age, and gender.

The following three measures were used together to achieve an appropriate correlation matrix of items to be used for exploratory factor analysis (Dziuban and Shirkey 1974):

1. Kaiser-Meyer-Olkin's (KMO) measure of sampling adequacy. Items were removed recursively from the data until a value of 0.68 was obtained, close to the guideline 0.7 recommended by Kaiser and Rice (1974).
2. Bartlett's (1950) test of sphericity. This had a significance level of less than 0.001 when guideline 1 had been achieved.
3. The anti-image correlation measure of sampling adequacy (MSA). Items with an MSA less than 0.5 were removed (Kaiser 1970).

As a result, eleven items were removed and are listed in Table 1. These items were interpreted as having little effect on the system of student retention, at least as far as this illustrative study is concerned, given the limited data set and the high level of retention from first year to second year.

Table 1. Items removed from the exploratory factor analysis.

Question No	Item
-	Age
-	Gender
-	Credits passed that are not a part of the programme of study
6	My possibility to continue with my studies is dependent on me working while I study
9	It is important for me to graduate at my University
13	I have achieved the study-results I expected during the first year
18	It is important for me to get a university degree
20	I have developed a good relationship with my teachers in the courses I have studied
26	First year physics courses have been inspiring
27	University physics courses are much different from my previous physics courses
28	First year physics courses have had a clear connection to everyday life

To decide the number of factors in the model we used a scree test (see scree plot in Figure 1). 'The scree test involves examining the graph of the eigenvalues ... and looking for the natural bend or break point in the data where the curve flattens out. The number of data points *above* the 'break' ... is usually the number of factors to retain' (Costello 2005, 3). To generate the scree plot, every item is treated as a vector that has an eigenvalue (length) of 1.0, before the optimizations of the sum of the vector projections on factors are carried out. For example, from Figure 1 we can see that an eigenvalue of seven leads to a One Factor solution which provides us with the information that all significant loadings in One Factor can be grouped, providing us with 7 times as much information as a single variable. It also means that the items in the factor share traits.

This led us to choose a Four Factor solution for the model (Hofstede 2001). The cut-off at Four Factors, and not Five (although they have nearly the same eigenvalue) was guided by seeing that a Five Factor solution provided one factor with only two variables that had significant loading (more than 0.32), which according to the analysis-method is

not appropriate for such a factor solution. The 'extra' factor would not give much more information than adding one or two other variables to the analysis or the questionnaire.

Figure 1. Scree plot showing eigenvalues and number of factors after varimax rotation. A Four Factor solution was chosen.

Significant item loadings for each factor were identified (See Table 2) by using a minimum loading of 0.32 on each item, which corresponds to a 10% shared variance between items (Tabachnick and Fidell 2001). Question 12 was retained at a loading of 0.313 (which is very close to 0.32).

The results of the exploratory factor analysis are shown in Table 2. Note that these results differ from the normal result in exploratory factor analysis where unique variables are sought for each factor. It is tempting to try to characterize the four factors (systems of components) in terms of the systems identified by others (such as university academic systems and social systems (Tinto 1975) and support systems (Bean 1980)) but this cannot be done because of the small sample size. What Table 2 does show is that there is an overlap of items between the four factors, each of which is a complex system in itself. This illustrates the complexity, the interdependence, and the nestedness of the system of student retention as a whole. It also highlights the existence of neighbour interactions between the four nested component systems, as well as the fact that they have fuzzy boundaries.

Our sample size is far too small to even attempt a tentative characterization of these four factors. A study using a sufficiently large enough sample size could characterize each factor according to the highest loading variables in that

factor. *To illustrate how this could be done*, we generated the following factors from our data: Factor 1; the status of the programme the students are studying; Factor 2; the sense of belonging; Factor 3; retention; Factor 4: financial issues, and has particularly fuzzy boundaries with Factors 1 and 3.

Table 2. Loading from the exploratory factor analysis giving the factors sorted adjacently by the number of shared items. Light grey shading denotes the items that have a loading above 0.32 in more than one factor.

Items	Factor 1	Factor 2	Factor 3	Factor 4
H.E. credits programme (HECwP)			0.542	
Retention			0.934	
Q1. Best university programme	0.788			
Q2. Family approval		0.472		
Q3. Satisfied with finances				0.836
Q4. Finances - focus on studies				0.833
Q5. Finances - teacher demands				0.796
Q7. Satisfied with curriculum	0.328	0.458		
Q8. Close friends encouragement		0.580		
Q10. I belong at my university		0.637	0.447	
Q11. Future employment	0.464	0.390		
Q12. My close friends rate this institution as high quality			0.313	
Q14. Satisfied with experience of higher education		0.687	0.411	
Q15. Easy to make new friends		0.842		
Q16. Right choice - university		0.683	0.399	
Q17. Right choice - programme	0.758			
Q19. It is important to get a degree from this programme	0.708			
Q21. Initiation weeks		0.855		
Q22. First year courses fit together				0.459
Q23. Previous knowledge		0.385		
Q24. Clear educational trajectory		0.447	0.396	
Q25. Faculty support		0.345	0.322	0.461
Q29. I intend to re-enroll	0.835			

Multidimensional scaling

Multidimensional scaling was used to visualize the network of components that influence student retention, and network theory data analysis tools and complexity thinking were used to interpret the results.

Network creation

The multidimensional scaling analysis used the same data as the exploratory factor analysis to determine the distances between items. A dimensional solution ranging from

two to four dimensions was explored, because multidimensional scaling usually provides a solution that has fewer dimensions than exploratory factor analysis on the same data (Schiffman et al. 1981).

To create a visualization of the network (Figure 2) we used the multidimensional proximities between items to identify items with relative closeness or proximity. Using the neighbour-interactions concept and an understanding of the structure of decentralized networks (Davis and Sumara 2006) from complexity thinking, we recursively lowered the cut-off for proximities, and network visualizations were produced. The statistical computing and graphics “R” program, together with the Statnet package (Handcock et al. 2003), was used for visualization and measurements. Iterations were run as long as the network continued to resemble a decentralized network, but were ended before the network broke down (after 15 iterations) and ceased to be connected (Freeman 1978). Two items were considered to be within each others’ ‘zone of influence’ when their proximity was below 0.25. The analysis was complete when the majority of the items had proximities less than 0.25. To retain the connectedness of the system Retention needed to have a cut-off of 0.5. HECOP, Gender, Q6 (studies dependent on working) and Q9 (importance of achieving a degree from this university) all dropped out at this cut-off level. These items are four of the eleven items that were dropped from the exploratory factor analysis. However, not all items that were dropped in the exploratory factor analysis were only loosely connected in the network. Moreover, three items (nodes) that may be particularly influential were identified. These three items were each present in two of the four factors in the exploratory factor analysis.

Note that as the cut-off level is lowered further, the system becomes less and less connected. At a cut-off proximity of 0.1 less than half the items remain connected to one another and there are very few connections between them compared to Figure 2.

Figure 2. Network visualization from multidimensional scaling analysis. Note that this visualisation shows just the connections, not the actual proximities. HECwP is higher education credits achieved within the programme.

Influential items

Influential items in the system are interpreted as items which are highly connected to other items, as measured by closeness and betweenness centrality. These items either have short paths to other items (closeness) or have a tendency to be part of a short path (betweenness). Thus information passes through these nodes (or items) frequently and they are key to effective and efficient information sharing. In a decentralized complex system these items contribute to the connectedness of the system and the removal of one of them could lead to the collapse of the system.

We used the closeness centrality and betweenness centrality scatter plot (Figure 3) to identify network items (nodes) that have a large influence in the network. Nodes with high closeness centrality and high betweenness centrality both distribute information effectively to a large proportion of the system, and are in a position of ‘control’ of other nodes’ influences on the system. Figure 3 shows nodes that are ‘close’ to other nodes and nodes which have a high frequency of being ‘between’ other nodes.

Consideration of Figure 3 shows that there are seven items with relatively high betweenness centrality as well as relatively high closeness centrality: Q12 (friends’ opinion of institutional quality), Q7 (satisfaction with one’s course curriculum), Q25 (faculty support), Age of the students, Q14 (students’ satisfaction of being at the

university), Q10 (the feeling of belonging at the university) and Q28 (physics is connected to everyday life). Item Q25 (faculty support) is interesting in that it seems to lie outside the broad band of points showing higher betweenness centrality vs higher closeness centrality: It has a much higher betweenness centrality than the rest of the items in the band. The same is also true for Q24 (clear educational trajectory). Connections between exploratory factor analysis results and multidimensional scaling results will be discussed below.

Figure 3. Closeness centrality and betweenness centrality scatter plot of the network created by the multidimensional scaling analysis proximities of items. All numbered markers correspond to same number question. Marker "reten" corresponds to the measurement of student retention. Marker named "HECwP" corresponds to students' Higher Education Credits achieved Within Programme. Marker "Age" corresponds to the age of the students.

Discussion

While not all aspects of a complex system such as student retention can be characterized by any research approach, we have tried to compellingly illustrate how bringing in a complexity thinking perspective can offer unique possibilities and interpretations. We have also attempted to show how complementary tools can be used in such studies. For example, our illustration drew on factor analysis to identify component complex systems within the broader complex system of student retention, and to demonstrate their nested and interdependent structure. In our example, four component systems emerged as a possible solution to the exploratory factor analysis and illustrated how a significant overlap, or fuzzy boundaries, between the systems may be interpreted in terms of the interdependence and interactions between systems and items within the systems that influence student retention. Items shared by multiple factors could be interpreted as key issues and may be areas that post secondary institutions may want to target with changes or efforts in order to improve student retention rates.

Multidimensional scaling was used to show the connectedness of items, to identify influential items and to visualize how the items of the complex system interact with one another. Complexity thinking provides new insights into student retention in terms of advocating for paying attention to items that have the potential to influence *the complex system as a whole*. This means that items should not be seen as direct linear influences, but mainly as *influencing* what takes place indirectly *through other items*. A stronger data set would allow for claims to be made about which items are more influential. Then targeted efforts to implement changes in the system of student retention could be made by manipulating particular (influential) items and the ensuing dynamics at several levels of analysis (or complex systems) could be observed and studied. With the identification of influential variables and a knowledge of the structure and dynamics of complex systems in general, targeted efforts could be made to positively influence the system as a whole.

Multidimensional scaling was chosen over other analysis techniques such as structural equation modelling and path analysis because these techniques rely on estimates of linear relations among a set of variables (Denison 1982). Multidimensional scaling techniques provide a non-linear approach to data analysis (Kaplan 2004). Using this idea we illustrated how a visual representation of the structure and dynamics of a complex system can be created.

What we were unable to illustrate with our current limited data set were issues relating to reliability and validity, however we feel that this has not detracted from our aim since the competent dealing with such issues can only be dealt with when an adequate data set has been obtained.

In conclusion, we believe that our illustrative analysis has managed to provide a credible example of a method of researching into complex issues in higher education which we believe has the potential to facilitate new, refreshing and exciting outcomes.

Acknowledgements

We thank the anonymous reviewers for their invaluable comments and suggestions that led to improving this article. We also thank Anne Linder and John Airey for their grammatical and logical comments, and Jannika Chronholm-Andersson for her developmental discussions, during the drafting of this article.

References

- Adams, W.K., K.K. Perkins, N.S. Podolefsky, M. Dubson, N.D. Finkelstein, and C.E. Wieman. 2006. New instrument for measuring student beliefs about physics and learning physics: The Colorado learning attitudes about science survey. *Physical Review Special Topics - Physics Education Research* 2, no.1: 1 – 14.
- Barnett, Ronald. 2007. *A will to learn: Being a student in an age of uncertainty*. Open University Press: New York.
- Bartlett, M.S. 1950. Tests of significance in factor analysis. *British Journal of Psychology* 3: 77 – 85.
- Bean, J.P. 1980. Dropouts and turnover: The synthesis and test of a causal model of student attrition. *Research in Higher Education* 12, no. 2: 155 – 187.
- Bean, J.P. 1982. Student attrition, intentions, and confidence: Interaction effects in a path model. *Research in Higher Education* 17, no. 4: 291 – 320.

- Bean, J.P. 2005. Nine themes of college student retention. In *College student retention: Formula for student success*, ed. A. Seidman, 215 – 244. Westport: Praeger Publishers.
- Bernhardsson, S. 2009. Structures in complex systems: Playing dice with networks and books. PhD diss., Umeå University.
- Booth, S., and Å. Ingerman. 2002. Making sense of physics in the first year of study. *Learning and Instruction* 12, no. 5: 493 – 507.
- Bourdieu, Pierre. 1984. *Distinction: A social critique of the judgement of taste*. Cambridge: Harvard University Press.
- Braxton, John M. 2000. *Reworking the student departure puzzle*. Nashville: Vanderbilt University Press.
- Braxton, J.M., and A.S. Hirshy. 2004. Reconceptualizing antecedents of social interactions in student departure. In *Retention and student success in higher education*, ed. M. Yorke and B. Londen, 89-102. Maidenhead: England: Open University Press.
- Braxton, J.M., and A. L. Lien. 2000. The viability of academic integration as a central construct in Tinto's interactionalist theory of college student departure. *Reworking the student departure puzzle*, ed. J. M. Braxton, 11-28. Nashville: Vanderbilt University Press.
- Brown, S.L., and K.M. Eisenhardt. 1997. The art of continuous change: Linking complexity theory and time-paced evolution in relentlessly shifting organizations. *Administrative Science Quarterly* 42, no. 1: 1 – 34.
- Cabrera, A.F., M.B. Castañeda, A. Nora, and D. Hengstler. 1992. The convergence between two theories of college persistence. *Journal of Higher Education* 63, no. 2: 143 – 164.
- Cabrera, A.F., A. Nora, and M.B. Castañeda. 1993. The role of finances in the persistence process: A structural model. *Research in Higher Education* 33, no. 5: 571 – 593.
- Cabrera, A.F., A. Nora, and M.B. Castañeda. 1993. College persistence: Structural equations modelling test of an integrated model of student retention. *Journal of Higher Education* 64, no. 2: 123 – 39.
- Committee on Science, Engineering, and Public Policy. 2007. *Rising above the gathering storm: energizing and employing America for a brighter economic future*. Washington, D.C.:The National Academies Press.
- Costello, A.B. 2005. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation* 10, no. 7: 1 – 9.
- Davis, Brent, and Dennis Sumara. 2006. *Complexity and education: Inquiries into learning, teaching and research*. New Jersey: Lawrence Erlbaum Associates.
- Davis, B., and E. Simmt. 2006. Mathematics-for-teaching: An ongoing investigation of the mathematics that teachers (need to) know. *Educational Studies in Mathematics* 61, 293 – 319.
- Dziuban, C.D., and E.C. Shirkey. 1974. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin* 81, no. 6: 358 – 361.
- Eaton, S., and Bean, J., 1995. An approach/avoidance behavioral model of college student attrition. *Research in Higher Education* 36, no. 6: 617 – 645.

- European Commission. 2004. *Europe needs more scientists: Report by the high level group on increasing human resources for science and technology*. Brussels: European Commission.
- Freeman, L.C. 1978. Centrality in social networks conceptual clarification. *Social Networks* 1, no. 3: 215 – 239.
- Gilstrap, D.L. 2005. Strange attractors and human interaction: Leading complex organizations through the use of metaphors. *Complicity: An International Journal of Complexity and Education* 2, no. 1: 55-69.
- Handcock, M.S., D.R. Hunter, C.T. Butts, S.M. Goodreau, and M. Morris. 2003. *Software tools for the statistical modeling of network data. Version 2.1-1*. Project home page at <http://statnet.org>, URL <http://CRAN.R-project.org/package=statnet>.
- Hofstede, Geert. 2001. *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations*. 2nd ed. Thousand Oaks, CA: Sage Publications.
- Hofstede, G., B. Neuijen, D.D. Ohayv, and G. Sanders. 1990. Measuring organizational cultures: A qualitative and quantitative study across twenty cases. *Administrative Science Quarterly* 35, no. 2: 286 – 316.
- Kaplan, David. 2004. *The SAGE Handbook of Quantitative Methodology for the Social Sciences*. California, Sage Publications.
- Kaiser, H.F., and J. Rice. 1974. Little jiffy, Mark IV. *Journal of Educational and Psychological Measurement* 34, no. 1: 111 – 117.
- Kaiser, H.F. 1970. A second generation Little jiffy. *Psychometrika* 35: 401 – 416.
- Kim, Jae-On, and Charles W. Mueller. 1978. *Factor analysis: Statistical methods and practical issues*. London: Sage Publications.
- Knoke, David, and Song Yang. 2008. *Social network analysis quantitative applications in the social sciences*. 2nd ed. California: Sage Publications.
- Lujan, H.L. and S.E. DiCarlo. 2006. Too much teaching, not enough learning: What is the solution? *Advances in Physiology Education* 30, no. 1: 17 – 22.
- Metz, G.W. 2004. Challenge and changes to Tinto's persistence theory: A historical review. *Journal of College Student Retention Research Theory and Practice* 6, no. 2: 191 – 207.
- Morrison, K. 2006. Complexity theory and education. Paper presented at the Asia-Pacific Educational Research Association International Conference, November, in Hong Kong.
- Mitchell, Margaret. 2009. *Complexity: A guided tour*. New York: Oxford University Press.
- Mowat, E., and B. Davis. 2010. Interpreting embodied mathematics using network theory: Implications for mathematics education. *Complicity: An International Journal of Complexity and Education* 7, no 1: 1 – 31.
- Pascarella, E.T., and P.T. Terenzini. 1979. Interaction effects in Spady's and Tinto's conceptual models of college dropout. *Sociology of Education* 52, 197-210.
- Paulsen, M.B. and E.P. St. John. 1997. The financial nexus between college choice and persistence. *New Directions for Institutional Research*, no. 95: 65 – 82.
- Redish, Edward F. 2003. *Teaching physics with the Physics Suite*. Hoboken, N.J.: Wiley.

- Schiffman, Susan S., Lance M. Reynolds, and Forrets W. Young. 1981. *Introduction to multidimensional scaling theory, methods, and applications*. Orlando: Academic Press, Inc.
- Schön, Donald A. 1983. *The reflective practitioner: How professionals think in action*. New York: Basic Books.
- Spady, W. 1970. Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange* 1, no. 1: 64 – 85.
- Spady, W. 1971. Dropouts from higher education: Toward an empirical model. *Interchange* 2, no. 3: 38 – 62.
- Statistics Sweden and National Agency for Higher Education. 2003. *Higher education. Students and graduated students in undergraduate education 2001/02*. Series: Utbildning och forskning [education and research] UF20 SM0301. SCB: Svante Öberg. Retrieved from: <http://www.scb.se/statistik/UF/UF0205/2000I02/UF20SM0301.pdf> (accessed 1st Dec, 2010)
- Statistics Sweden and National Agency for Higher Education. 2005. *Higher Education. Students and graduated students in undergraduate education 2003/04*. Series: Utbildning och forskning [education and research] UF20 SM0501. SCB: Svante Öberg. Retrieved from: <http://www.hsv.se/download/18.1db453f212664e532507ffe293/UF20SM0501.pdf> (accessed 1st Dec, 2010)
- Statistics Sweden and National Agency for Higher Education. 2007. *Higher education. Students and graduated students in undergraduate education 2005/2006*. Series: Utbildning och forskning [education and research] UF20 SM0701. SCB: Kjell Jansson. Retrieved from: <http://www.hsv.se/download/18.1db453f212664e532507ffe297/UF20SM0701.pdf> (accessed 1st Dec, 2010)
- Statistics Sweden and National Agency for Higher Education Higher Education. 2009. *Students and graduated students at first and second cycle studies 2007/08*. Series: Utbildning och forskning [education and research] UF20 SM0902. SCB: Mats Vadman. Retrieved from: <http://www.hsv.se/download/18.1db453f212664e532507ffe302/UF20SM0902.pdf> (accessed 1st Dec, 2010)
- Statistics Sweden and National Agency for Higher Education. 2010. *Higher education. Students and graduated students at first and second cycle studies 2008/09*. Series: Utbildning och forskning [education and research] UF20 SM1002. SCB: Stefan Lundgren. Retrieved from: http://www.scb.se/Statistik/UF/UF0205/2008L09D/UF0205_2008L09D_SM_UF20SM1002.pdf (accessed 1st Dec, 2010)
- Summerskill, J. 1962. Dropouts from college. In *The American college: a psychological and social interpretation of higher learning*, ed. N. Sanford, 627-657. New York: Wiley.
- Tabachnick, Barbara G., and Linda S. Fidell. 2001. *Using multivariate statistics*. Boston: Allyn and Bacon.
- Terenzini, P.T., and E.T. Pascarella. 1980. Toward the validation of Tinto's model of college student attrition: A review of recent studies. *Research in Higher Education* 12, no. 3: 271 – 282.

- Tinto, V. 1975. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research* 45, no. 1: 89 – 125.
- Tinto, Vincent. 1987. *Leaving college: Rethinking the causes and cures of student attrition*. Chicago: University of Chicago Press.
- Tinto, V. 1997. Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education* 68, no. 6: 599 – 623.
- Tinto, V. 2010. From theory to action: Exploring the institutional conditions for student retention. *Higher Education: Handbook of Theory and Research*, no. 25: 51 – 89.
- Waldrop, M. Mitchell. 1992. *Complexity: The emerging science at the edge of order and chaos*. New York: Simon & Schuster.
- Yorke, Mantz, and Bernard Longden. 2004. *Retention and student success in higher education*. Maidenhead: Open University Press.

Appendix 1: Questionnaire

The questionnaire is given in Table A1, together with the sources/motivations for each of the questions. It was constructed after extensive discussions with colleagues and senior students, to ensure that it was fitting for the Swedish higher education context. Fifteen questions were taken from the eighteen identified by Cabrera et al. (1992) as having high explanatory value for student retention. Three questions were omitted: the one about financial support was covered by three financial questions that were added (Questions 4-6); the question about transferring to other universities was omitted because it is not common in the Swedish context; and the third question about developing personal relationships was covered by Question 21 about the initiation weeks programme. Ten of the fifteen questions were modified slightly to fit the Swedish context; all fifteen questions were translated into Swedish and then independently translated back into English for a meaning-matching check and final translation adjustments.

Since Cabrera et al.'s survey was performed in 1992, other work was used to add fourteen questions to the questionnaire. Questions 1 and 17 were added to address students' programme of study. Questions 4 – 6 were added about students' financial attitudes. Questions 20, 23, and 25 were added to better explore the importance of teacher-student interactions. Question 21 was added to account for the tradition of initiation weeks that students have during their first weeks at the university. Questions 22 and 24 were added to explore how the students experience their courses as 'pieces of a whole' (Booth and Ingerman 2002). Questions 26 – 28 were included to ground the questionnaire in our context of undergraduate engineering and physics programmes.

Table A1. Questions asked and their source(s) and/or motivation (changes/new questions)

Question	Source/motivation
Q1. I am studying one of the best programmes at the university.	In Sweden, students in general choose "programmes" that are planned to lead towards a degree
Q2. My family approves of my attending my University.	Cabrera et al. (1992)
Q3. I am satisfied with my financial situation.	Cabrera et al. (1992), mod: In Sweden, University Education is free and students receive study loans and grants from the government towards living expenses.
Q4. My financial situation allows me to focus on my studies as much as I want.	Expansion of Question 3 - see Cabrera et al. 1993; Paulsen and St. John 1997
Q5. My financial situation allows me to focus on my studies as much as the teachers demands.	Expansion of Question 3 - see Cabrera et al. 1993; Paulsen and St. John 1997

Q6. My possibility to continue with my studies is dependent on me working while I study.	Expansion of Question 3 - see Cabrera et al. 1993; Paulsen and St. John 1997
Q7. I am satisfied with my course curriculum.	Cabrera et al. (1992)
Q8. My close friends encourage me to continue attending my University.	Cabrera et al. (1992)
Q9. It is very important for me to graduate at my University.	Cabrera et al. (1992) simplified
Q10. I feel I belong at my University.	Cabrera et al. (1992)
Q11. My degree at this university will help me secure future employment.	Cabrera et al. (1992)
Q12. My close friends rate this university as a high quality institution.	Cabrera et al. (1992)
Q13. I have achieved the study-results I expected during the first year.	Cabrera et al. (1992) modified
Q14. I am satisfied with my experience of H.E.	Cabrera et al. (1992) modified
Q15. It has been easy for me to meet and make friends with other students at this university.	Cabrera et al. (1992)
Q16. I am confident I made the right decision in choosing to attend at my university.	Cabrera et al. (1992)
Q17. I was right when choosing to study this programme.	New question
Q18. It is important for me to get a university degree.	Cabrera et al. (1992), terminology change
Q19. It is important for me to get a degree from this particular programme.	Cabrera et al. (1992), terminology change
Q20. I have developed a good relationship with my teachers in the courses I have studied.	See Spady 1970;1971; Tinto 1975;1987;1997, Terenzini and Pascarella 1980
Q21. The initiation weeks were a good start for my program studies.	New question to account for the traditional "introduction period" that students have during their first weeks at the university.
Q22. It is clear to me how the courses during the first year fit together.	See Booth and Ingerman 2002

Q23. The teaching has corresponded well with my previous knowledge.	See Spady 1970;1971; Tinto 1975;1987;1997, Terenzini and Pascarella 1980
Q24. My educational trajectory is clear for me.	See Booth and Ingerman 2002
Q25. Faculty staff have provided me with the support I needed to succeed in my studies.	See Spady 1970;1971; Tinto 1975;1987;1997, Terenzini and Pascarella 1980
Q26. First year physics courses have been inspiring.	See Lujan and DiCarlo 2006
Q27. University physics courses are much different from my previous physics courses.	See Tinto 1975; 1987; 1997
Q28. First year physics courses have had a clear connection to everyday life.	See Adams et al. 2006; Redish 2003
Q29. I will re-enrol at this programme of study next autumn.	New question