Understanding processes of identity development and career transitions
van der Gaag, Mandy
Developmental trajectories leading up to student drop-out

Mandy A.E. van der Gaag
E. Saskia Kunnen
Paul L.C. van Geert
ABSTRACT

Due to governmental incentives, preventing student drop-out has become a top priority for many universities over the past few years. We aim to shed light on the processes preceding dropout so that universities may have the possibility to identify at-risk students at an early stage. In our intensive longitudinal study, we have distinguished two types of educational commitment and exploration trajectories among first year students of psychology: a trajectory that is characterized by certainty (strong commitments and little exploration, both being stable over time) and a trajectory that is characterized by high levels of doubt (weak commitments and much exploration, both fluctuate much over time). Individuals who show trajectories characterized by doubt are at an elevated risk of dropping out of higher education. These typologies are useful to detect dropout at an early stage as they are relatively easy to detect, but the prediction of dropout can be much improved by gaining insight into more sophisticated features of the individual trajectories. In our study, dropout is most accurately predicted if individuals show a decreasing trend of commitment over time combined with an increasing trend in the amount of fluctuations in exploration, or if they show an increasing trend in their level of exploration. We have shown that trajectories of exploration and commitment seem to be relevant predictors to university dropout. Influencing these processes of exploration and commitment may have high gains with relatively little costs. By showing ways to identify at risk students early, and suggesting a few avenues through which processes of exploration and commitment can be guided, we hope to have provided evidence so that universities may try new, individual and process-oriented ways to tackle dropout.
INTRODUCTION

University drop-out is a common problem, with large consequences for both the individual and society. For example in Europe 20% to 50% of all students who start higher education do not finish it (Quinn, 2013). University dropouts have a high chance to be without higher education, training and employment for a long time (European Training Foundation, 2014). This can have long term consequences: these individuals are more likely to experience personal tragedies, to come into contact with the justice system, and cause significant welfare costs for society (e.g., Coles, Godfrey, Kueng, Parrott, & Bradshaw, 2010). To prevent dropout and the long term negative consequences this may have, it is key to understand the process of student drop-out and to identify students at risk of dropping out at an early stage. In this paper we aim to contribute to this knowledge by studying within-individual trajectories of educational commitment and exploration throughout the first year of higher education and investigating how these trajectories are related to dropout.

Identifying Students At Risk For Drop-Out

Many universities have sought to reduce the chances of dropout by identifying at-risk students, but as the high attrition numbers illustrate this is by no means an easy task. A straightforward and increasingly widely adopted strategy is to prevent individuals who have a high risk of dropping out of higher education to enter the university in the first place. In practice however, this approach encounters a number of obstacles. First of all, it is not ethically sound to select students at the gate using some of the strongest predictors of academic success: socio-economic status (Jury, Smeding, & Darnon, 2015), gender, or migratory status (e.g., Stoessel, Ihme, Barbarino, Fisseler, & Stürmer, 2015). An alternative strategy is to use less controversial predictors to select students, like personality or motivational factors (e.g., Richardson, Abraham, & Bond, 2012). This strategy encounters another practical obstacle: it is difficult to assess prospective students accurately on these variables, as these students are tempted to present themselves in such a way that will improve their chances to get admitted (e.g., Niessen, Meijer, & Tendeiro, 2017). Universities are then left with what is perhaps the most common strategy of selection: selecting on past performance, usually in the form of the grades obtained during high school, as these grades are good predictors of student performance (Westrick, Le Robbins, Radunzel, & Schmidt, 2015). However, this strategy also encounters an obstacle: it may be overly restrictive, excluding potentially talented individuals, as many high
school students are notorious underachievers (e.g., 28% in a study of Vaes, Gilar, Miñano, & Castejón, 2016).

Early identification of students who have a high risk of dropping out during the first year of higher education has three advantages over selection at the gate: 1) it is not troubled by the obstacles described above, 2) predictors of dropout that are measured during the first year of higher education, as opposed to predictors measured before the start of higher education, are found to be the strongest predictors of dropout and 3) it becomes possible to design interventions to help these at-risk students to obtain their degree. Many studies demonstrate the value of predictors measured during the first year of higher education, but these studies differ in the type of predictor they focus on. A recent meta-analytical study of Westrick, Le, Robbins, Radunzel, and Schmidt (2015) investigated the effect of high school grades, the scores on a standardized test for high school performance (ACT), and first year academic performance on student retention. They found that first year academic performance is the strongest predictor of student retention. Another meta-analytical study by Robbins et al. (2004), showed that psychosocial characteristics directly related to the academic context were the strongest predictors of student retention, particularly academic self-efficacy, academic goals, and academic-related skills. Other studies have found that the fit with educational context as experienced by the students is the best predictor of student retention. Examples of such fit factors are the commitment students feel towards their educational institution (Strom, 1985), the certainty students feel about their choice of educational trajectory (Metzner, Lauer, & Rajecki, 2003; Neuville et al., 2007), the experienced congruence between interest and the chosen major (Allen & Robbins, 2010), and perceiving an academic fit (Elffers, Oort, & Karsten, 2012; Wintre, Bowers, Gordner, & Lange, 2006). Among so many studies showing different variables that contribute to dropout it is hard to determine the most useful predictor – which variable should a university monitor in order to identify at-risk individuals? As we shall demonstrate below, this question may be answered by placing these variables in the context of theories on the individual process of dropout.

Processes of Dropout

Tinto’s classic model of student attrition (1975, 2012) is perhaps the most influential longitudinal process theory on student attrition. The model combines several important predictors in one theory. In this theory, a student arrives at university with a certain set of attributes (prior schooling, family background, skills and abilities). The
student forms educational commitments (i.e., commitments towards the institution and towards her own educational goals). These commitments affect the type of experiences the student has within the academic and social system, such as peer-group interaction and academic performance. These experiences subsequently affect social and academic integration, which then feeds back into the educational commitments of the student. Ultimately, these educational commitments inform the decision to depart. This model implies that although psychosocial (Robbins et al., 2004) and academic performance factors (Westrick et al., 2015) may have an important predictive value on whether students dropout, their effect is exerted through a mediating, more proximal role of commitment. As such, the educational commitment of an individual is perhaps a useful indicator to monitor regularly in order to identify possible dropouts at an early stage.

Commitments are also of central importance in many classic and modern theories on identity (e.g., Bosma & Kunnen, 2001; Erikson, 1956; Marcia, 1966). From an identity perspective, commitments theoretically develop through active exploration – such as reconsiderations and in-depth explorations of the current commitment (Crocetti, Rubini, & Meeus, 2008; Luyckx, Goossens, Soenens, & Beyers, 2006; Marcia, 1966). Commitments can be seen as the bond that is felt with a certain context, while exploration is the behavior through which one can investigate and re-evaluate this bond. Commitment and exploration are both core factors of identity development and are tightly connected - explorations have been shown to be related to commitment dynamics on both a micro level (e.g., Klimstra et al., 2010) and on a macro level (e.g., Luyckx, Goossens, & Soenens, 2006). It is not hard to imagine how these processes may affect each other. An individual who feels only a weak connection with her chosen educational trajectory (a low commitment) may start to reconsider whether this trajectory truly fits her (an act of exploration). Such explorations may lead the individual to conclude that the educational trajectory is indeed not that fitting at all, further diminishing the commitment she feels towards her educational trajectory. The combined processes of commitment and explorations have been shown to be related to dropout (e.g., Luyckx, Goossens, & Soenens, 2006). Thus as a complementing force to commitment, theories on identity development predict that exploration could also be an important factor to monitor if we are to effectively identify individuals in processes headed for dropout.

The levels of these educational commitments and the amount of explorations are not necessarily stable throughout the academic year. If we apply Tinto’s model to a micro-level process – i.e., an individual student in his day-to-day life – we can
get an idea of how fluctuations may occur. Imagine for example, that on one day a student receives a large compliment from a fellow student (i.e., positive peer group interaction). This experience fortifies the individuals’ feeling of belonging with his fellow students (i.e., increases social integration) and reinforces his feeling that this is indeed the right place for him (i.e., fortifies educational commitment). However, the next week, he may receive a bad grade (i.e., negative academic performance) which undermines his feeling of belonging at university (i.e., decreases academic integration) and weakens the feeling that this educational trajectory is right for him (i.e., weakens educational commitment). Indeed, when framing Tinto’s model in such a micro-level process perspective, it seems plausible that commitment levels may go up and down frequently, on a weekly or even on a daily basis, and levels of exploration may move with it. This has in fact been shown to be true for micro-level educational commitments and explorations from an identity development perspective (Klimstra et al., 2010).

In systems science, fluctuations are considered important information. They give information about the internal dynamics and stability of a system. Fluctuations can be precursors of a transition to a qualitatively different state in any kind of system (see for example Scheffer et al., 2012). This idea has been applied in psychology as well, for example therapeutic contexts, where an increase in fluctuations of emotions is an indicator of qualitative change (Lichtwarck-Aschoff, Hasselman, Cox, Pepler, & Granic, 2012). It has also been applied in the study of development, particularly in young children, where an increase in fluctuations may signal a transition to a new developmental phase (for an overview, see Van Dijk & Van Geert, 2015). If we consider dropout as a transition to a qualitatively different state, it is plausible that also this transition is preceded by fluctuations in the students’ trajectories of commitment and exploration. Thus it seems important to take the amount of and changes in fluctuations in commitment and exploration into account when trying to identify individuals in processes headed for dropout.

**Present study**

In the previous section we have argued that if the many predictors of dropout are integrated in theories on the individual process of dropout, the investigation of trajectories of commitment and exploration seems particularly fruitful so that we may ultimately detect potential dropouts in an early stage. In the present study we explore the types of commitment and exploration trajectories that may exist among first year students and we relate these types of trajectories to the probability
that the students drop out of higher education. Moreover, we investigate which of these trajectory characteristics predict dropout the best. Thus we aim to answer three questions: 1) Which types of trajectories of exploration and commitment can be distinguished? 2) How are such trajectories related to dropout? 3) Which characteristics of the commitment and exploration trajectories are particularly important for predicting dropout?

There are many forms of commitment and exploration one could focus on. In this study we have chosen to particularly focus on the strength of the commitment that an individual feels towards her chosen, specific educational trajectory, and to what extent she explores whether this trajectory fits her. The choice for these particular forms of commitment and exploration is related to the context in which we study educational commitment processes: higher education in Europe, particularly in the Netherlands, where all bachelor programs focus on specific topics of study (e.g. psychology, history, ecology, etc.). Students have to choose this topic of study before they start the bachelor. Thus the choice process of students, the subsequent commitment they form and the explorations they may perform, is strongly focused on a specific educational trajectory. Such a conceptualization is slightly different from the concept of commitment as proposed in Tinto’s model, where commitments are formed to personal goals and the educational institution, but it is in line with forms of commitment and exploration commonly studied in identity research (e.g., Klimstra et al., 2010; Van der Gaag, De Ruiter, & Kunnen, 2016).

We use a process approach in order to identify students headed for dropout at an early stage. This means that we study exploration and commitment both within individuals and as processes over time. Our micro-level longitudinal study consists of very frequent measurements over the course of the first year of higher education. This allows us to not only study the level of exploration and commitment, but also study the gradual changes in these factors over time, as well as their fluctuations and changes in these fluctuations.

**METHOD**

**Participants**
Our sample consists of 115 first year bachelor students who have chosen to pursue the educational trajectory of psychology at a university in the north of the Netherlands. The mean age of this group was 19.3 ($SD = 1.8$) at the beginning of the study.
The majority of participants is female (81%, $N = 93$; versus 19%, $N = 22$ male), this is in line with the gender distributions (predominantly female) within this particular educational trajectory (psychology). The students participated as part of their curriculum – they are required to gather credits for research participation. They can freely choose the type of research in which to participate. All participants are Dutch speaking and live in the northern part of the Netherlands. Of these 115 participants, 12 dropped out of higher education after completing their participation in this study (specifically, they stopped pursuing their chosen educational trajectory anywhere between the end of the first academic year and the start of their third academic year), and 103 persisted in their educational trajectory (specifically, they continued their specific educational trajectory at least until the start of the third academic year).

The participants filled in weekly reports throughout a large part of their first academic year. Eighteen individuals (13%) were excluded from the original sample ($N = 134$) because they completed less than 80% of the required amount of experiences reports. In addition, one participant was excluded because she misunderstood the instructions. This makes a total of 19 excluded individuals, leaving 115 individuals in our total sample.

The amount of experience reports is different for two subsamples of the total sample: a ‘long’ subsample where 30 weekly experience reports were asked of the students, and a ‘short’ subsample where only 22 experience reports were asked. The ‘short’ subsample is shorter due to practical constraints – as multiple researchers make use of the same pool of research participants, we were limited in the amount of participant time that we could use. The included participants of the long subsample ($N = 71$) completed 29 experience reports on average ($SD = 2.0$). The included participants of the short subsample ($N = 44$) completed 22 experience reports on average ($SD = 0.9$). We have no reason to expect systematic differences between the two subsamples: they differ in the amount of weeks spent in this study, but the measured variables and population are the same. We have therefore taken them together in our main analysis.

**Procedure**

We collected data weekly throughout three quarters of the first academic year for the long subsample, and throughout slightly more than half an academic year for the short subsample. For the long subsample the data collection started in November, and continued until June, for a total period of seven months. For the short
subsample data collection started in January and continued until June, for a total period of five months. The participants in both subsamples were asked to fill out the same online questionnaire every week. This questionnaire contained a qualitative and quantitative section; for this study we only use the quantitative measures of exploration and commitment.

To reduce the chance of attrition over this long period of data collection, participants were allowed to choose the moment in the week to fill out the questionnaire that suited them best. This did not have to be the same moment each week. They were also allowed to skip two weeks during the data collection period (but not right after each other). Because of the substantial sustained effort required of the participants, the students were rewarded accordingly, with an attractive amount of credits.

The data of the long subsample was collected in three cohorts: first year students from academic years 2011–2012 \( (N = 12) \), 2012–2013 \( (N = 25) \) and 2013–2014 \( (N = 34) \). The data of the short subsample was collected only in academic year 2013–2014 \( (N = 44) \). Data on dropout was obtained through the administrative office of the university, all individuals in this study agreed to share this information.

**Measures**

The participants filled out a weekly online questionnaire, the Repeated Exploration and Commitment Scale in the domain of Education (RECS-E; Van der Gaag, et al., 2016). In our analysis we included the micro-level exploration and commitment measures of the RECS-E that were administered among all cohorts: one measure of exploration (exploration of fit: “Have you asked yourself whether this educational trajectory is right for you?”) and one measure of commitment (commitment to choice: “Do you stand by your choice for this particular educational trajectory?”). Both were rated on a Likert scale of 1 (not at all) to 6 (very much).

**Analysis**

We have performed our analysis in three steps corresponding to our three research questions. First, we used a cluster analysis to classify the different types of commitment and exploration trajectories. Second, we compared the cluster memberships of students who drop out with those who persist – in this way we determined whether the chance to dropout is different for the different types of trajectories. Third, we have employed a simple machine learning technique – generating decision trees – to determine which of the trajectory characteristics are particularly important for predicting dropout, and which are less important.
1) Cluster analysis of individual trajectories

We have performed a cluster analysis on several characteristics of the individual trajectories of exploration and commitment. We have included a total of eight variables in the cluster analysis – four characteristics of the exploration trajectory, and four characteristics of the commitment trajectory. We used the time-serial data of each individual to determine these trajectory characteristics. Particularly, we have calculated each individuals’ average level of exploration and commitment, the general trend of change in these levels, the amount of fluctuations in both variables, and the general trend of change in these fluctuations. The average level is calculated by taking the average score of commitment and of exploration using the time-serial data of each individual. The variability of the individual trajectories is the average absolute change in commitment and exploration from one week to another. For example, if an individual first scored 5 on commitment, then 3, and then 4, the change scores are -2 and 1, the absolute change scores are thus 2 and 1, and the variability in commitment is the average of these absolute change scores: 1.5.

In individual time series, such a variability measure is considered superior to using standard deviation as measure of variation (Kunnen, 2012). The general linear trend is calculated by subjecting the time-serial commitment and exploration data of each individual to a linear regression analysis. This analysis results in the best fitting linear equation of which we extract the slope to represent the general trend of change (e.g., upwards, downwards, or leveled) of commitment and exploration over time. We have performed the same procedure to calculate the general linear trend of the variability - using the weekly absolute change scores as input for the linear regression.

We have standardized all eight of the variables to equate their impact on the clustering solution. Variables with larger ranges and variances may have a larger influence on the clustering solution (Henry, Tolan, & Gorman-Smith, 2005). Standardization is necessary when the theoretical ‘weight’ of each of the variables is considered equal (Milligan, 1996). As we indeed consider our variables of equal importance we have standardized them by transforming the variables into z-scores (i.e., \( Z_x = (X – Mean_x) / SD_x \)). We have used the classic and widely known K-means algorithm (Lloyd, 1982) to perform the cluster analysis.

K-means aims to minimize dissimilarity between individuals within a cluster. The k-means algorithm first chooses the cluster centers randomly, then assigns the individuals that are closest to the cluster center to the one cluster. When all individuals are assigned to clusters, the cluster centers are calculated again and the process repeats for a certain amount of iterations – we have used 100 iterations – until the cluster
centers are optimized. Because the cluster centers are initially chosen randomly, slight variations may occur in each resulting clustering solution (i.e., the initialization problem; Tan, Steinbach, & Kumar, 2006). We have employed a common solution to this (Tan et al., 2006): we have performed the cluster analysis several times (50) and from the resulting clustering solutions we have chosen the optimal one – the solution with the least variance within each cluster and the most variance between the clusters (i.e., the solution with the smallest number when subtracting the between cluster sum of squared error from the total within cluster sum of squared error).

K-means requires that the amount of clusters (k) is defined a priori. We have chosen to divide our data over two clusters: this is the simplest partitioning possible which we consider appropriate considering our modest sample size, and a two cluster solution will allow straightforward comparison with the probability of dropping out of higher education in the second step of our analysis. To perform the cluster analysis we have used the package cluster (Maechler, Rousseeuw, Struyf, Hubert, & Hornik, 2016) in R (R Core Team, 2016). Finally, we perform t-tests to compare the two clusters on the means of the 8 trajectory variables that we have included in the analysis. In this way we discover which of the 8 variables are truly important for distinguishing the groups.

2) Comparing trajectories of dropouts and persisters

In the second step of our analysis we have investigated whether students who drop out tend to have different types of commitment and exploration trajectories than students who persist. To do so, we show the distribution of cluster membership among dropouts and among persisters. We have performed a one-sided Fisher’s exact test (Fisher, 1922) in R (R Core Team, 2016) to compare the distributions, which is particularly suitable when distributions contain conditions with small numbers (this is the case in our data, as we shall show in the results). The resulting $p$-value represents the total probability that the null hypothesis – that dropouts and persisters are equally likely to show a certain type of trajectory – is true if we observe distributions as extreme or more extreme than the one we found.

3) Decision trees

In the third and final step of our analysis, we ‘grow’ a decision tree to predict dropout as accurately as possible (without overfitting) based on our 8 trajectory variables. It is a relatively simple machine learning technique that generates easily interpretable results (for an introduction see Tan, et al., 2006). A decision tree consists of nodes and branches. The top (‘root’) node represents the most important variable used for
classification (i.e., the variable that separates the data the best). Each branch represents a cut-off score, for example an average commitment score higher than 5. Each individual satisfying this cut-off score is transported to the node below. This new node may be another classification variable (an ‘internal node’) or it may be a ‘terminal’ node – a final classification in the class ‘dropout’ or ‘persist’. Many algorithms can be used to grow a tree. We have chosen to use the widely known algorithm C4.5 (Quinlan, 1998) which has been chosen as the number one algorithm in data mining by a large panel of data-mining experts (Wu et al., 2008).

In order to perform well, a decision tree needs to be provided with classification groups of approximately equal size. However, in our data we do not have equally sized classification groups (103 individuals that persist and 12 individuals that drop out). This is a common problem when predicting anomalies in a dataset. A commonly used (and well performing) solution is to create synthetic examples of the minority class in a procedure called SMOTE (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). This SMOTE procedure entails that the three nearest neighbors to each data point are connected with lines, and synthetic examples of the minority class are added randomly somewhere along these lines. This synthetic data is generated solely for training purposes – these simulated individuals are removed once the tree is trained and the trees’ accuracy is of course only tested on the empirical data. A downside to using simulations to train a tree is that the generated trees can vary in structure and cut-off scores – this makes sense as there is randomness in the synthetic data that is generated in the SMOTE procedure. Thus the next step was to choose a good classification tree among many possible trees.

As a general rule, highly complex trees (i.e., trees with many nodes and branches) are likely to lead to overfitting – the tree may fit the sample very well, but it may not be generalizable to the population, and it is not easy to interpret (Tan et al., 2006). Contrastingly, very simple trees are likely to lead to underfitting – even though it is easily interpretable, it may make many errors in classifying the individuals in the sample. The challenge is to find a balance between overfitting and underfitting.¹

We have employed this general rule to choose a simple tree among the trees that we have generated. In our choice process we mainly focused on preventing overfitting because overfitting endangers generalizability which is already an issue.

¹ Optimally, a tree is grown using a training dataset and then tested on a new dataset. This requires a large amount of data in all categories. We only have few (12) individuals in the dropout category and therefore we have decided to take an inductive approach: only train the tree but take elaborate measures to prevent overfitting.
because of the relatively small dataset. To this end, we have made sure that each node in the tree would contain at least 10 individuals in the training process – this way each node would not be too specific for only a small set of individuals. In addition, we have employed a process called ‘post-pruning’ (see also Tan et al., 2006), which is basically an algorithm that removes branches of the tree that add little to the correct classification of individuals. We generated 50 trees in this way. These trees varied with respect to their complexity, ranging from two to five branches deep. In our first selection step, we only selected the relatively simple trees - those with a maximum depth of 3 (i.e., the maximum length of the chained branches connecting the root node to a terminal node never exceed 3). This left 34 trees in our selection set. We then excluded two trees because the same variables reoccurred in different sections of the tree, and such trees are hard to interpret (Tan et al., 2006).

To create a balance between overfitting and underfitting we selected two trees from the 32 trees left in our set: a relatively complex tree with high accuracy (i.e., with a chance of overfitting) and a relatively simple tree with lower accuracy (i.e., with a chance of underfitting). The relatively complex tree was selected by picking the tree with the highest average accuracy in correctly classifying the dropouts and the persisters among the 32 trees left in our dataset. The simple tree was selected by first taking a subsample of only the most simply structured trees – containing only two variables and with a maximum depth of two. This set contained 11 simple trees and from these we selected the tree with the highest accuracy in correctly classifying the empirical data. By showing both decision trees we gain insight into the relative importance of each of the predictor variables.

RESULTS

1) Cluster Analysis of Individual Trajectories

The descriptive statistics of the two clusters solution are presented in Table 1, these numbers are illustrated in the individual time serial trajectories of commitment and exploration in Figure 1. The individual trajectories in the first cluster (N = 64, see Table 2 for an overview of frequencies) are characterized by certainty. The commitment trajectories of individuals in this first cluster tend to show high levels (M = 5.4, SD = 0.5), a leveled linear slope over time (M = 0.00, SD = 0.02), little variability in the level of commitment (M = 0.3, SD = 0.2) and a slightly declining linear slope of this variability (M = -0.01, SD = 0.02). The exploration trajectories in this first cluster tend
Table 1 Summary statistics of each of the trajectory variables separately for each cluster. The two clusters have been compared on each trajectory variable using a t-test. The p-value of this t-test is reported below the corresponding trajectory variable on the bottom row.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Commitment Level</th>
<th>Commitment Slope of Level Variability</th>
<th>Commitment Slope of Variability</th>
<th>Exploration Level</th>
<th>Exploration Slope of Level Variability</th>
<th>Exploration Slope of Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
</tr>
<tr>
<td>Cluster 1 'Certain'</td>
<td>5.4 (0.5)</td>
<td>0.00 (0.02)</td>
<td>0.3 (0.2)</td>
<td>-0.01 (0.02)</td>
<td>1.7 (0.6)</td>
<td>-0.02 (0.05)</td>
</tr>
<tr>
<td>Cluster 2 'Doubting'</td>
<td>4.2 (0.9)</td>
<td>-0.01 (0.04)</td>
<td>0.6 (0.2)</td>
<td>-0.03 (0.04)</td>
<td>3.0 (0.8)</td>
<td>-0.02 (0.07)</td>
</tr>
<tr>
<td>P-value of difference</td>
<td>&lt; 0.001</td>
<td>0.009</td>
<td>&lt; 0.001</td>
<td>0.168</td>
<td>0.473</td>
<td>0.110</td>
</tr>
</tbody>
</table>

Figure 1 Illustration of the individual time-serial trajectories of both commitment (left figures) and exploration (right figures) separate for cluster 1 ('certain' top figures) and cluster 2 ('doubting' bottom figures). The colored lines represent the trajectories of different individuals (these are loess smoothed scores for clear illustration).

to show low levels \((M = 1.7, SD = 0.6)\), a declining linear slope over time \((M = -0.02, SD = 0.05)\), little variability in the level of exploration \((M = 0.6, SD = 0.3)\) and no linear change in this this variability \((M = 0.00, SD = 0.02)\).

The individuals in the second cluster \((N = 51)\) show very different types of trajectories, characterized by frequent doubt. The commitment trajectories of individuals within this cluster tend to have relatively low levels \((M = 4.2, SD = 0.9)\), a slightly declin-
ing linear slope over time ($M = -0.01$, $SD = 0.04$), much variability ($M = 0.6$, $SD = 0.2$), but also a linear decrease of this variability ($M = -0.03$, $SD = 0.04$). The exploration trajectories of individuals in this cluster are characterized by high levels ($M = 3.0$, $SD = 0.8$), a declining linear slope over time ($M = -0.02$, $SD = 0.07$), much variability ($M = 1.1$, $SD = 0.4$), and a slight linear decrease in this variability ($M = -0.01$, $SD = 0.04$).

Not all trajectory variables differ significantly between the clusters (see also table 1). Individuals in the two clusters show large mean differences in their level of commitment ($\Delta M = 1.2$, $p < 0.001$), their level of exploration ($\Delta M = -1.3$, $p < 0.001$), their variability in commitment ($\Delta M = -0.3$, $p < 0.001$), and their variability in exploration ($\Delta M = -0.5$, $p < 0.001$). They also show a small but insignificant difference in their mean slope of commitment ($\Delta M = -0.01$, $p = 0.009$). We find no significant differences between individuals in the two clusters in their slope of exploration ($\Delta M = 0.00$, $p = 0.473$), in their slope of the variability in exploration ($\Delta M = -0.01$, $p = 0.110$), nor in their slope of the variability in commitment ($\Delta M = -0.02$, $p = 0.168$).

2) Comparing Trajectories of Dropouts and Persisters

Figure 2 shows the number of students that persist ($N = 103$) versus the amount of students that drop out ($N = 12$), divided over the two clustering solutions (see also table 2). The distribution of these two groups over the two clusters is unequal: Fischer exact test demonstrates a low probability of finding a distribution such as this or more extreme ($p = 0.025$) if dropouts and persisters were equally divided over.

![Figure 2](image.png)

**Figure 2** Distribution of individuals that persist (left two bars) and those that drop out (right two bars) over the clusters obtained from our analysis of time-serial trajectories: certain trajectory (blue) and doubting trajectory (yellow).
the clusters. Of the 103 students that persist, a small majority of individuals (59\%) shows a trajectory characterized by certainty, while a smaller group of individuals (41\%) shows a trajectory characterized by frequent doubt. This pattern is reversed for the 12 students that drop out. Most of these dropouts show a trajectory characterized by doubt (75\%), while only few show a trajectory characterized by certainty (25\%). Thus individuals in the doubting cluster run a higher risk of dropping out than individuals in the certain cluster. But with an overall accuracy of 61\%, this prediction is far from perfect. Individuals who show commitment and exploration trajectories characterized by certainty are predicted to persist rather accurately: 95\% of these individuals are correctly predicted to persist\(^2\). However, individuals who show trajectories characterized certainly do not always drop out: only 18\% of these individuals are correctly predicted to dropout.

### 3) Decision Trees

We have generated two decision trees to provide a more refined prediction of dropout. It also provides insight into the type of trajectory markers, and in the exact decision criteria, that perform best in distinguishing the dropouts from the persisting students.

#### Simple tree

The simplest decision tree that we have generated correctly classifies 80\% (92 out of 115) of the individuals. This simple tree performs particularly well in identifying

---

\(^2\) But given the highly unequal distribution (90\% persisters and 10\% dropouts), there is already a high chance of being accurate when predicting that any given individual will persist based chance alone (i.e., 90\% accuracy)
dropouts: 92% ($N = 11$) of the dropouts are correctly classified. The tree performs less well in identifying persisters: 79% ($N = 81$) of the persisters are correctly classified, thus leaving a relatively large portion of individuals (21%; $N = 22$) who are wrongly predicted to be a dropout. The tree includes only 2 predictors: the slope of commitment and the slope of exploration. This simple tree, including its predicted classes and cut-off criteria, is illustrated in Figure 3.

In this simple tree, the slope of commitment is a main predictor of dropout: 19 individuals are predicted to dropout (Figure 3, box A.) because they show a relatively strong decreasing linear trend in their commitment levels over time ($M = -0.06; SD = 0.02$). 37% of these individuals are accurately classified as a dropout: 7

![Figure 3](image_url)

**Figure 3** Simple decision tree that predicts membership of the classes ‘Dropout’ (boxes A. and C., in red) and ‘Persist’ (box B., in green). Each class contains the total amount of individuals classified in it (indicated with $N =$ ...) and the portion of correctly classified individuals. This correctly classified portion is illustrated with a miniature pie chart: the portion of the pie that corresponds to the color of the surrounding box is correctly classified, contrasting colors represent misclassification (e.g., a green pie section within a red box is the portion of persisters that are wrongly classified as dropouts). The cut-off criteria are presented next to each branch (i.e., an arrow that connects two nodes) as standardized scores – e.g., $\leq -0.81$ means that an individual must score 0.81 standard deviations below the average to be classified in the class below.
of the 19 individuals who meet the criterion (standardized slope of commitment ≤ -0.81) have truly dropped out.

The slope of commitment needs to be combined with the slope of exploration to predict the large class of individuals that are likely to persist \((N = 82, \text{ Figure } 3, \text{ box B.})\). This class consists of individuals who tend to show a levelled or slightly increasing linear trend in their commitment scores \((M = 0.01; SD = 0.02)\) and who also show a decreasing linear trend in their exploration scores \((M = -0.04; SD = 0.05)\). Individuals in this class are predicted to persist with an accuracy of 99\%: 81 of the 82 individuals who meet this criterion (standardized slope of commitment ≤ -0.81 and standardized slope of exploration ≤ 0.94) have truly persisted.

Finally, the combination of the slope of commitment and the slope of exploration is also needed to classify a small group of individuals as dropouts \((N = 14, \text{ Figure } 3, \text{ box C.})\). Similar to the persisting class, these individuals tend to show a levelled linear trend in their commitment scores over time \((M = 0.00; SD = 0.02)\), but contrary to the persisting class they tend to show an increasing linear trend in their exploration levels over time \((M = 0.07; SD = 0.03)\). 29\% of the individuals in this class are correctly classified as a dropout: 4 of the 14 individuals who meet the criterion (standardized slope of commitment > -0.81 and standardized slope of exploration > 0.94) have truly dropped out.

**Complex tree**

The more complex tree is more accurate overall: 91\% \((N = 105)\) of the individuals have been classified correctly. Compared to the more simple tree described above, this more complex tree has a bit more trouble with correctly classifying dropouts: 83\% \((N = 10)\) of the dropouts are classified correctly (thus leaving 2 dropouts classified wrongly as a persister, compared to only 1 misclassification of this type by the simple tree). The complex tree performs markedly better in classifying persisters than the simple tree: 92\% \((N = 95)\) of the persisters are correctly classified (thus only a small portion of the persisters are misclassified as dropouts: 8\%; \(N = 8\)).

The more complex tree is similar to the simple tree in two ways. First, the most important criterion for distinguishing dropouts from persisters is also in the complex tree the slope of commitment – it even predicts the exact same cut-off criteria as the simple tree does. Second, the slope of exploration also occurs as an important predictor in the complex tree as it does in the simple tree, but with slightly different cut-off criteria. The complex tree is primarily different from the simple tree because
of the addition of two new predictors: the slope of variability in commitment and the slope of variability in exploration (see also Fig. 4).

The slope of commitment is combined with the slope of variability in exploration to classify individuals as likely to persist (Figure 4, box A.). This persisting class consists of 9 individuals who tend to show a decreasing trend in their commitment levels \((M = -0.06; SD = 0.02)\) and also tend to show a decreasing trend in the variability of their exploration \((M = -0.05; SD = 0.03)\) – i.e., they show less or smaller ups and downs in exploration as time progresses. 89% of the individuals in this class

---

**Figure 4** Complex decision tree that predicts membership of the classes ‘Dropout’ (boxes B., C., and E., in red) and ‘Persist’ (boxes A., and E., in green). Each class contains the total amount of individuals classified in it (indicated with \(N\)) and the portion of correctly classified individuals. This correctly classified portion is illustrated with a miniature pie chart: the portion of the pie that corresponds to the color of the surrounding box is correctly classified, contrasting colors represent misclassification (e.g., a green pie section within a red box is the portion of persisters that are wrongly classified as dropouts). The cut-off criteria are presented next to each branch (i.e., an arrow that connects two nodes) as standardized scores – e.g., \(\leq -0.81\) means that an individual must score 0.81 standard deviations below the average to be classified in the class below.
are correctly classified as persisting: 8 of the 9 individuals who meet the criterion (standardized slope of commitment ≤ -0.81 and standardized slope of the variability in exploration ≤ 0.24) have truly persisted.

These same two variables, but with different values for the slope of variability in exploration, are used to classify individuals as likely to drop out (Figure 4, box B.). This dropout class consist of 10 individuals who tend to show a decreasing trend in their commitment levels similar to the persisting class described above ($M = -0.05; SD = 0.02$), but contrary to this persisting class they tend to show a linear increase in the variability of their exploration ($M = 0.02; SD = 0.03$) – i.e., they show more or stronger ups and down in exploration as time progresses. Interesting to note is that this linear increase in fluctuations of exploration is not systematically accompanied by a linear increase in the level of exploration ($M = 0.00; SD = 0.06$). 60% of the individuals in this dropout class are correctly classified: 6 of the 10 individuals who meet the criterion (standardized slope of commitment ≤ -0.81 and standardized slope of the variability in exploration > 0.24) have truly dropped out.

Similar to the simple tree, the combination of the slope of commitment and the slope of exploration classifies a small group of individuals ($N = 5$) as dropouts (Figure 4, box C.). These individuals tend to show a levelled linear trend in their commitment scores over time ($M = 0.00; SD = 0.02$) and tend to show an increasing linear trend in their exploration levels over time ($M = 0.09; SD = 0.02$). Interesting to note is that this increase in exploration seems to be gradual – individuals in this class tend to show a levelled or decreasing amount of variability in exploration over time ($M = -0.01; SD = 0.01$). 60% of the individuals in this class are correctly classified as a dropout: 3 of the 5 individuals who meet the criterion (standardized slope of commitment > -0.81 and standardized slope of exploration > 1.60) have truly dropped out.

Another small dropout class ($N = 3$) is predicted by adding the slope of the variability in commitment to the slope of exploration and commitment (Figure 4, box D.). Individuals in this dropout class tend to show a slightly increasing trend in their commitment scores over time ($M = 0.02; SD = 0.04$), and the variability in these commitment scores decreases linearly over time ($M = -0.07; SD = 0.01$). There seem to be large differences in this small group in the linear trend of the exploration levels, it is levelled or slightly increasing on average ($M = 0.01; SD = 0.05$). 33% of the individuals in this class are correctly classified as a dropout: only 1 individual out of the 3 individuals who meet the criterion (standardized slope of commitment > -0.81, standardized slope of exploration ≤ 1.60, standardized slope of the variability...
in commitment ≤ -1.93) has truly dropped out, thus calling into question the usefulness of this criterion.

Finally, a large class ofpersisters ($N = 88$) is predicted based on the same three variables as the small dropout class described above – the slope of exploration, the slope of commitment and the slope of the variability in commitment – but with different values for the slope of the variability in commitment (Figure 3, box E.). This persisting class consists of individuals who tend to show a levelled linear trend in their commitment scores ($M = 0.00; SD = 0.02$), the variability in these commitment scores tends to stay levelled or decreases slightly over time ($M = -0.01; SD = 0.02$), and their exploration scores tend to decrease ($M = -0.03; SD = 0.05$). Individuals in this class are predicted to persist with an accuracy of 99%: 87 of the 88 individuals who meet the criterion (standardized slope of commitment > -0.81, standardized slope of exploration ≤ 1.60, standardized slope of the variability in commitment > -1.93) have truly persisted.

**DISCUSSION**

**Main Findings**

The cluster analysis shows us that individuals can be distinguished from one another based on their time-serial trajectories of exploration and commitment. Five features of the individual trajectories seem to distinguish the individuals the best: the level and variability of both commitment and exploration, and to a lesser extent also the linear change in commitment over time – these features differ significantly between the two clusters. The three other trajectory markers that we included in our analysis do not distinguish well between the two clusters of individuals (the slope of exploration, and the slope of variability in exploration and the slope of variability in commitment do not differ significantly between the two clusters). In general, the difference between the individuals in the two clusters can be characterized by certainty in the one cluster, and doubt in the other. Individuals in the cluster characterized by certainty tend to show a stable, high level of commitment that shows no particular increasing or decreasing trend over time, and a low level of exploration that does not fluctuate much. Individuals in the cluster characterized by doubt tend to show a highly fluctuating, low level of commitment that decreases slightly over time and a high level of exploration that fluctuates much.
The two types of trajectories are related to the chance that students drop out of higher education: students who persist in their chosen educational trajectory more often show a trajectory characterized by certainty, while students who drop out more often show a trajectory characterized by doubt. When viewed from a different angle – an angle perhaps more useful to universities – individuals who show trajectories characterized by certainty are rather reliably predicted to persist in their studies. However, individuals who show trajectories characterized by doubt do not necessarily dropout - the large majority of students who show such a doubting trajectory do not dropout. Thus having a doubting trajectory seems to be a necessary, but not a sufficient condition for dropout.

We have refined this dropout prediction by generating two decision trees. The simplest of these two trees generates a straightforward prediction of two types of individuals who are likely to dropout: individuals who show a decreasing trend in their level of commitment over time, and individuals for whom this trend in commitment is leveled but who do show an increase in their level of exploration. Contrastingly, individuals who are likely to persist tend to show a decreasing trend in their level of exploration, and a slight increasing trend in their level of commitment. These criteria predict dropout rather well, but are still quite likely to also falsely classify students who persist as dropouts. The second, more complex decision tree that we have generated performs better in this regard than the simple tree and is more accurate overall, with the downside of having more complex decision criteria. The complex decision criteria are particularly useful to better classify individuals who show a decreasing trend in their commitment. While the simple tree predicts that all these individuals will dropout, the more complex tree predicts that whether this is true depends on changes in the variability of exploration: individuals with a decreasing commitment are only likely to drop out if their amount of exploration starts to vary more over time (i.e., exploration shows more or stronger ups and downs), but if this is not the case then they are likely to persist. The complex tree adds another prediction: individuals with a leveled or increasing trend in both commitment and exploration may drop out if their level of commitment stabilizes over time. This prediction only accurately identified one dropout, so the general usefulness of this criterion is questionable.
Types of Micro-level Educational Commitment and Exploration Processes

It indeed seems useful to monitor the level of educational commitment and exploration regularly in order to identify students at risk for dropout. Tinto’s model (1975, 2012) of student attrition suggests that commitment is a proximal factor in student attrition, and that these dynamics of commitment are part of a larger process - experiences within and outside the educational setting may shape educational commitment, and this commitment may in turn affect academic and social experiences. Our study verifies that such commitment dynamics seem to exist on a micro-level and adds that exploration may also be an important part of this process, as has also been shown to be the case in studies that are focused on identity development (Crocetti et al., 2008; Klimstra et al., 2010; Luyckx et al., 2006). Indeed, we have found that micro-level dynamics in educational commitment and exploration on a week-to-week basis is common, but also that students differ in the type of dynamics they tend to show.

As is reflected in the results of our cluster analysis, both the level of and fluctuations in exploration and commitment play an important role in characterizing individual trajectories. Having a high level of commitment and a low level of exploration seems to be a comfortable, steady state for an individual. In this state, both exploration and commitment do not fluctuate much, meaning that the individual is most of the time fairly certain of her educational choice, and she experiences few periods of only mild doubt. The other state an individual may be in is much more volatile. When commitment is low, it is not stably low – it fluctuates. This fluctuation means the individual experiences frequent periods of little certainty in her educational choice, alternated with periods of relative certainty. This unstable state of commitment is accompanied by a generally elevated level of exploration, which also fluctuates much. This means that the individual experiences frequent periods of intense exploration, actively doubting and reconsidering the fit of her chosen educational trajectory, alternated with more quiet periods.

This co-occurrence of fluctuations with different levels of exploration and commitment is an interesting addition to the identity literature. In this field, fluctuations are not commonly included in typologies of individual differences in commitment and exploration: these typologies are commonly based on the level of exploration and commitment (e.g., Meeus, van de Schoot, Keijsers, Schwartz, & Branje, 2010), sometimes also including the general change trend (e.g., Luyckx, Schwartz, Goossens, Soenens, & Beyers, 2008). However, our results suggest that these levels of
commitment and exploration are intertwined with their stability: high levels of commitment tend to be stable while low levels of commitment tend to fluctuate much. Conversely, high levels of exploration tend to fluctuate much while low levels of exploration tend to be stable. Thus identity classes may not only be characterized by the levels of exploration and commitment, but also by the stability in these levels. Inclusion of these characteristics may lead to the distinction of new or other identity classes, and a new perspective on how identity develops. Thus an important step in future research on identity development is to investigate not only general individual tendencies of the level of exploration and commitment and gradual changes in these levels, but also stability characteristics.

**Predicting Dropout with Trajectories of Exploration and Commitment**

We have shown that it is indeed useful to investigate trajectories of exploration and commitment for the early identification of individuals at risk for dropout and that many different types of information on individual trajectories can be used for the prediction of dropout. The accuracy of the predictions depends on which of the features of the trajectory is used for the prediction. Overall, it is interesting to note that the best predictions of dropout (generated by the decision trees) make little use of the features of trajectories that come forward as the most different between individuals (generated by the cluster analysis). The trajectory features that differentiate well between individuals are the level of exploration and commitment, and the amount of variability in this level: these features differ the most between the two types of individual trajectories that we have characterized as doubting and certain. These groups of students are perhaps also relatively easy to distinguish for teachers and mentors at universities: the students’ trajectories seem to be markedly different in ways that make intuitive sense. These categories also do a moderately good job at distinguishing dropouts from persisters – they work particularly well for identifying students that will persist and they filter out the large majority of dropouts, but they also tend to falsely flag many individuals as a risk for dropout while they will actually persist in their studies.

Subtler trajectory characteristics perform better at predicting dropout, but they are perhaps less visible – and may be harder to detect for teachers or mentors at universities – as they pertain to gradual change trends that cover long periods of time. All three of the best predictors of dropout that we have discovered in this study pertain to such change trends: a decreasing trend in the level of commitment, an increasing trend in the level of exploration and an increasing trend in the amount
of fluctuations in exploration. These general trends are quite difficult to identify among the many fluctuations that are also present in many individual trajectories. Such general trends can only be separated from the fluctuations within individuals (that are also abundantly present in a large group of individuals in our study) when individuals are followed with a high frequency over a long period of time.

That we find fluctuations to be a distinguishing feature of trajectories with an elevated risk for dropout makes sense from a general system science perspective (Scheffer et al., 2012; Van Dijk et al., 2015). From this perspective, we inferred that fluctuations may be a precursor of a qualitative change in a system – in our case this qualitative change is dropout. However, it is not true that all individuals showing a trajectory characterized by fluctuations (the ‘doubting’ types) also drop out of higher education: many persisters also show this type of trajectory. Thus for some individuals, having such fluctuations is apparently not a problem, at least not to such an extent that they drop out of higher education – perhaps to some individuals, frequent doubting is a common and stable state.

We also found that more frequent or more intense periods of exploration over time is an important predictor of dropout. This makes sense from a systems perspective: an increase in fluctuations in a system may precede a transition to a qualitatively different state (Van Dijk et al., 2015). Curiously, we find the opposite result for commitment: an increase in stability in commitment is identified as a predictor of dropout. However, this criterion has only accurately classified one individual as a dropout and misclassified two persisters as a dropout, thus it is highly questionable whether this criterion is truly useful to identify dropouts or if the one individual for whom this criterion worked well is somehow exceptional. Perhaps some external event or circumstance forced this individual to quit. But then again, it could also be an indicator of a more serious dilemma in using commitment as a predictor of dropout: perhaps the dropout event itself – as in, the actual unregistering from the educational trajectory - is not the transition to a qualitatively different state that we should be looking at. It may very well be that an individual has made up her mind long before she formally unregisters and that this is what we see when her level of commitment stabilizes, which points to the need to carefully consider how transitions should be defined.

Recommendations for Practice

Overall, our study shows that in order to identify university dropouts in an early stage, it is useful to look at individual trajectories of educational exploration and
commitment over time. We have shown that this can be done with obvious and easily observable criteria, such as characterizing individuals as certain or doubting. In our study – among psychology students – we find that students have a high chance of persisting in their studies if they are fairly certain of their educational choice most of the time and if they experience only few periods of mild doubt. Contrasting, we find that students have an elevated chance of dropping out if they frequently doubt whether the educational trajectory is truly fitting and generally experience little certainty in their educational choice. This classification as ‘certain’ or ‘doubting’ will filter out manypersisters and most of the dropouts, but has the disadvantage that it tends to falsely classify many doubters: the majority of doubters will actually persist in their studies. Thus using these classes as predictors has the upside that they can be identified relatively easily in practice – e.g., by taking a only a few measures among students or even by qualitative estimations of a teacher or mentor – but has the downside that many will be falsely identified as dropout-risk. This may lead to many resources being spent unnecessarily on the supervision of students who are not truly at risk of dropping out.

We have found that better predictions can be obtained by focusing the analysis of individual trajectories on more sophisticated features that require more frequent measurement in order to be estimated correctly. In our study, very useful predictors of dropout are a general decreasing trend of commitment and an increasing trend of exploration. We obtained an even more refined prediction by adding increases of fluctuation in exploration as a predictor: an increase in ups and down of exploration seems to distinguish the dropouts from the persisters among the group of individuals who all show a decreasing trend in their commitment. Using these predictors has the upside that it will lead to better predictions, but has the downside that the predictors require frequent measurement over a long period of time in order to be estimated correctly.

Thus it seems that universities need to balance their need to identify possible dropouts early with practical constraints on obtaining the detailed trajectory information that is needed to make the most accurate predictions. The choice seems to come down to this: better predictions cost more resources in measuring the trajectories, but will probably cost less resources in student guidance – as the predictions will be more accurate in identifying the students that will dropout, guidance can be deployed more efficiently. In order to use the best and most sophisticated predictors, a university needs to have a system in place that allows for refined and frequent measurement of exploration and commitment. In this is not feasible,
relatively simple measures already go a long way in identifying risk groups – i.e., the level of exploration and commitment, and the amount of fluctuations in these levels – with the downside that a rather large group of students are likely to be flagged to be at risk.

Generally, there are at least two advantages for universities when they monitor processes of exploration and commitment to try and identity individuals at risk for dropout in an early stage. First, such processes have the potential to be influenced by universities in a direct way, with relatively little cost. This contrasts other measures of early identification of possible dropouts, such as psychosocial characteristics of the students like self-esteem (Robbins et al., 2004). Such characteristics may be much harder to change for a university, requiring much more investment and expertise on the part of the teachers that is also perhaps not part of the core business of a university. Other important predictors of dropout such as academic performance (Westrick et al., 2015) are probably already monitored and heavily invested in, leaving little to be gained by investing more. This brings us to the second advantage: influencing these processes of exploration and commitment may have high gains with relatively little costs.

One can imagine relatively simple interventions that support students in their explorations and commitments. Universities can for example organize talks in mentoring groups about how students feel about their studies thus allowing doubting students to feel validated and reassured in their doubts. Or universities can for example implement structured exploration as part of assignments, allowing the students to come more quickly to an answer to the question of whether the educational trajectory truly fits them. Indeed, if universities try to help doubting students to find the answers they are looking for they may facilitate some students to pursue their studies with more confidence, while they may facilitate other students to decide in an early stage that this is not the right path for them.

Limitations
An important limitation of this study is the small and specific sample of first year students – most are female, all are from a particular University in the Netherlands, and they have all chosen psychology as the topic of their bachelor studies. Of course, the results need to be replicated in other contexts, and among a more diverse population, before we can infer the generalizability of our results. The small sample with few dropouts makes the predictions we make specific to the dropouts and persisters that happen to be included in our study. This means that it is highly pos-
 possibile that some idiosyncrasies may be at the heart of some of the predictive criteria. Our predictions are thus very much in need of replication among other populations, because student populations can be very different depending on the major they have chosen (Slijper, 2017). Our results are thus best viewed as hypotheses that need further testing. Nevertheless, our study provides insight into which variables seem promising for detecting dropout early, and provides some innovative ways to do this, which we have demonstrated to indeed be useful in predicting dropout.

We have only few dropouts in our sample compared to European statistics (Quinn, 2013). This may be due to the specific sample and the timing of our study may also play a role. To get valid trajectory measures we have included individuals who completed at least 80% of the measurements in our study. This means that they were in this study until roughly the end of the academic year, and students that have dropped out earlier in the year were not included. Expanding the sample by replicating the study might help to catch more of these dropouts. Another solution is to start measuring earlier in the academic year, so that valid trajectory measures can still be obtained, while also including dropouts that quit earlier in the year. This also means that our predictive trajectory measures can only be used to detect dropout among individuals that persist in their educational trajectory for at least a couple of months. Students that dropout in the first couple of days or weeks can probably not be detected with this method.

Another important limitation concerns the cluster analysis. Clustering techniques are sensitive to the particular technique used and choices made. If we would have used other clustering techniques or chosen another amount of clusters to classify the individuals in, we might have arrived at a different clustering solution. Moreover, other studies of trajectories of exploration and commitment (e.g., Luyckx et al., 2008; Meeus et al., 2010) included several types of exploration and commitment, while we have limited ourselves to one type of commitment and one type of exploration in our cluster analysis. If more types of exploration and commitment are included, perhaps trajectories can be characterized differently. Thus the two cluster solution presented here should not be interpreted as the only or most accurate reflection of reality. It is best seen as a useful way to categorize reality, as it allowed us to gain insight in the types of trajectories of commitment and exploration that may be related to dropout.

Lastly, we have not cross validated the decision trees. For more generalizable predictions the best procedure is to train the tree on a small subset of data and train it on another. We felt our set was too small for this procedure and have used all data
to train the tree while taking other measures to prevent over fitting, such as pruning the tree and selecting simple trees (see also the method section).

**Conclusion**

Studying individual micro-level processes of educational commitment and exploration seems to offer useful information to identify individuals at risk for dropout in an early stage, at least in this study with a quite specific sample of first year university students majoring in psychology. As we have shown, highly accurate early identification using processes of exploration and commitment is possible, but whether this is practically feasible depends on the availability of facilities to frequently measure exploration and commitment over a long period of time. If this is not present, perhaps early identification using simpler typologies as we have also presented is the more feasible option which is also rather effective in detecting dropout, but bears the cost of also identifying many false positives. By showing these ways to identify at risk students early, and suggesting a few avenues through which processes of exploration and commitment can be guided, we hope to have inspired researchers and universities to try a new, more individual and process-oriented way to detect and tackle dropout.
REFERENCES


European Training Foundation (2014). *Young people not in employment, education or training in the EU neighbourhood countries*. Turin, Italy: European Training foundation.


