ONDERWIJSVORMEN EN WERKLOOSHEID

Dockx J. & De Fraine B.

ONDERWIJSVORMEN EN WERKLOOSHEID

Dockx J.& De Fraine B.

Promotor: B. De Fraine

Research paper SONO/2018.OL1.1/09

Gent, januari 2018

Het Steunpunt Onderwijsonderzoek is een samenwerkingsverband van UGent, KU Leuven, VUB, UA en ArteveldeHogeschool.

Gelieve naar deze publicatie te verwijzen als volgt:
Dockx, J. & De Fraine, B. (2018). Onderwijsvormen en werkloosheid. Steunpunt Onderwijsonderzoek, Gent.

Voor meer informatie over deze publicatie
jonas.dockx@kuleuven.be; info@lisoproject.be

Deze publicatie kwam tot stand met de steun van de Vlaamse Gemeenschap, Ministerie voor Onderwijs en Vorming.

p.a. Coördinatie Steunpunt Onderwijsonderzoek
UGent – Vakgroep Onderwijskunde
Henri Dunantlaan 2, BE 9000 Gent

Deze publicatie is ook beschikbaar via www.steunpuntsono.be en www.lisoproject.be

Beleidssamenvatting

In het Vlaamse secundair onderwijs zijn er vier onderwijsvormen: het algemeen secundair onderwijs (aso), het technisch secundair onderwijs (tso), het beroepssecundair onderwijs (bso) en het kunstsecundair onderwijs (kso). Binnen het aso wordt daarbij vaak een onderscheid gemaakt tussen klassieke talen en moderne studierichtingen. Deze onderwijsvormen worden pas formeel ingericht vanaf de tweede graad van het secundair onderwijs, maar in de praktijk spreken leerlingen, ouders en scholen al in termen van onderwijsvormen in de eerste graad. In heel wat scholen zijn de onderwijsvormen reeds te herkennen in het onderwijsaanbod van de eerste graad. In het tweede leerjaar van de eerste graad worden namelijk basisopties ingericht die aansluiten op deze onderwijsvormen. De meeste scholen gebruiken hun pedagogische vrijheid voor het invullen van lesuren in het eerste leerjaar ook als voorbereiding op de onderwijsvormen in de bovenbouw. In de eerste graad bereiden het eerste leerjaar B en het beroepsvoorbereidend leerjaar voor op het bso.

Bij beleidsmakers is er discussie over mogelijke effecten van deze onderwijsvormen op de ontwikkeling van vaardigheden voor het functioneren op de arbeidsmarkt. Voorstanders argumenteren dat onderwijsvormen die aansluiten op de vaardigheden en interesses van leerlingen de ontwikkeling van gespecialiseerde vaardigheden mogelijk maakt. Tegenstanders argumenteren echter dat sociale ongelijkheid tussen leerlingen versterkt wordt doordat de vaardigheden in de hoger gepercipieerde onderwijsvormen voordeliger zijn op de arbeidsmarkt.

In wetenschappelijk onderzoek worden onderwijsvormen *tracks* genoemd. Er zijn diverse studies die onderzoeken hoe de *track* de verdere loopbaan van voormalige studenten beïnvloedt. De meeste van deze studies tonen dat een beroepsgerichte *track* volgen gemiddeld voordelig is bij de overgang naar de arbeidsmarkt. Deze leerlingen hebben meer kans op werk. Dit voordeel verdwijnt echter na verloop van tijd en sommige studies vinden dat een beroepsgerichte *track* zelfs nadelig wordt op lange termijn. Verder toont landenvergelijkend onderzoek dat de beroepsgerichte *tracks* enkel voordelig zijn bij de overgang naar de arbeidsmarkt wanneer deze een sterke geïntegreerde stage hebben.

Onderzoek naar de effecten van *tracks* op de langetermijnloopbaan van leerlingen in Vlaanderen is vereist om na te gaan of deze de ongelijkheid tussen leerlingen versterken. Als *tracks* namelijk deze ongelijkheid versterken, dan verwachten we dat *tracks* met een instroom van initieel sterker presterende leerlingen ook meer kans hebben op werk. Hiervoor moet de gemiddelde kans op werk per *track* vergeleken worden. Eventuele verschillen in kans op werk tussen *tracks* zijn dan wel mogelijk toe te schrijven aan verschillen in instroom van leerlingen. Daarom moet ook onderzocht worden of er effecten zijn van *tracks* op vergelijkbare leerlingen die in verschillende *tracks* zitten. Verder onderzoeken we of de effecten van *tracks* veranderen over de loopbaan heen.

Er zijn dus drie onderzoeksvragen:

1. Verschillen *tracks* in gemiddelde kans op werk?
2. Verschillen *tracks* in gemiddelde kans op werk voor vergelijkbare leerlingen?

3. Veranderen de verschillen tussen *tracks* in gemiddelde kans op werk (voor vergelijkbare leerlingen) over de loopbaan heen?

Voor dit onderzoek gebruiken we de gegevens van het onderzoek ‘Longitudinaal Onderzoek Secundair Onderwijs’ (LOSO). De substeekproef bestaat uit 4333 leerlingen die in september 1990 startten in het secundair onderwijs in 57 Vlaamse scholen. We onderscheiden vier groepen van studiekeuzes in het eerste jaar secundair onderwijs: (1) klassieke talen (KT), (2) moderne wetenschappen (MW), (3) technisch onderwijs (TO) en (4) beroepsvoorbereidend onderwijs (BV). Hoewel er in het eerste jaar secundair onderwijs nog geen officiële onderwijsvormen onderscheiden worden, sluit de studiekeuze in het eerste jaar SO wel sterk aan bij de onderwijsvormen die in de bovenbouw zullen volgen. In dit Engelstalige rapport wordt daarom wél gesproken over ‘tracking’ in het eerste jaar secundair onderwijs, omdat het gaat over het groeperen van leerlingen voor een volledig schooljaar voor (quasi) alle vakken.

De steekproef is verspreid over de vier ‘*tracks*’ in 1990 als volgt: 1522 leerlingen zaten in KT, 1084 leerlingen zaten in MW, 1047 leerlingen zaten in TO en 680 leerlingen zaten in BV. In onze analyses onderscheiden we leerlingen die in dezelfde *track* bleven en leerlingen die van *track* veranderden. 461 leerlingen van KT bleven in hun *track* terwijl 1061 leerlingen veranderden. 445 leerlingen van MW bleven in hun *track* terwijl 639 leerlingen veranderden. 493 leerlingen van TO bleven in hun *track* terwijl 554 leerlingen veranderden. De LOSO-scholen die kozen voor een heterogene klassamenstelling in het eerste jaar, werden geschrapt uit de steekproef van deze studie omdat er dus niet aan *tracking* wordt gedaan. Toetsen en vragenlijsten afgenomen aan de start van het secundair onderwijs (september 1990) gelden als startmeting op basis waarvan we gelijkaardige leerlingen in verschillende tracks vinden. Om de kans op werkloosheid te onderzoeken werd de maandelijkse kans op werkloosheid onderzocht van januari 1995 tot december 2015. Dit werd vervolgens gemodelleerd aan de hand van twee kansen: de kans om werkloos te worden wanneer men actief was (studeren of werken) en de kans om actief te worden wanneer men werkloos was.

Om vergelijkbare leerlingen in verschillende *tracks* te vinden gebruikten we *matching* methoden. Deze zijn gericht op het vinden van vergelijkbare personen in verschillende omgevingen. Leerlingen werden *gematched* op basis van schoolse prestaties, sociaaleconomische achtergrond en psychosociale variabelen die gemeten waren in september 1990. In totaal werd de vergelijkbaarheid van de leerlingen bepaald aan de hand van 14 variabelen. Er bleken (voldoende) vergelijkbare leerlingen waren tussen bepaalde *tracks*, namelijk de *tracks* die in gepercipieerde hiërarchie opeenvolgend zijn. KT wordt daarom vergeleken met het MW, MW wordt vergeleken met TO en TO wordt vergeleken met BO. De verschillen tussen *tracks* op de kans om werkloos te worden en de kans om actief te worden berekenen we tweemaal: (1) zonder het *matchen*, dus voor alle leerlingen, en (2) na het *matchen* van vergelijkbare leerlingen in verschillende *tracks*.

Voor de eerste onderzoeksvraag vinden we dat de gepercipieerde hiërarchie tussen de tracks zich ook uit in de kans om werkloos te worden wanneer men actief is. We stellen vast dat de voormalige leerlingen van KT de laagste kans hebben op werkloos worden. Daarna volgen de voormalige leerlingen van MW, de voormalige leerlingen van TO en de voormalige leerlingen van BO. Voor de kans om actief te worden wanneer men werkloos is wordt stellen we vast dat de voormalige leerlingen van KT de hoogste kans hebben om actief te worden. Daarna volgen de voormalige leerlingen van MW, de voormalige leerlingen van TO en de voormalige leerlingen van BO.

Voor de tweede onderzoeksvraag vinden we voor vergelijkbare leerlingen in verschillende *tracks* dat er in *tracks* met een gemiddeld sterkere leerlinginstroom significant minder kans hebben om werkloos te worden. We zien echter dat bij de vergelijking KT en MW, en bij de vergelijking MW en TO het effect klein is. Enkel voor de vergelijking TO en BV is het effect groot. Voor de kans om actief te worden wanneer men werkloos is vinden we enkel een significant positief effect voor MW in vergelijking met TO. Bij de andere vergelijkingen is er geen significant verschil.

Voor de derde onderzoeksvraag vinden we geen indicatie dat er een voordeel is voor leerlingen van het TO en BO vroeg in de loopbaan na het secundair onderwijs. Deze bevinding geldt zowel voor de tracks in hun geheel als wanneer we enkel naar vergelijkbare leerlingen kijken.

We onderzochten ook of starten in een hoger gepercipieerde *track* en dan veranderen naar een lager gepercipieerde *track* verschilt van starten in een lager gepercipieerde *track*. De resultaten tonen dat dit geen merkbaar verschil veroorzaakt voor de kans op werkloosheid. De *track* waarin men het secundair onderwijs beëindigd, niet diegene waarin men start, is van tel.

Een sterk punt van dit onderzoek is dat door de matching-methode nagegaan kan worden hoe vergelijkbare leerlingen zouden presteren als ze in een andere *track* zouden zitten. Dit is vooral mogelijk doordat *tracking* in Vlaanderen een eigenschap heeft die niet kenmerkend is voor de meeste andere onderwijssystemen. In Vlaanderen verloopt het verdelen van leerlingen in *tracks* immers niet op basis van objectieve criteria (bijvoorbeeld een instaptoets). Hierdoor verschillen de *tracks* wel gemiddeld op het vlak van instroomniveau, maar vinden we nog steeds veel vergelijkbare leerlingen terug in verschillende *tracks*. In andere onderwijssystemen zien we dat er minder of nauwelijks vergelijkbare leerlingen zijn in verschillende tracks.

We concluderen dat de *track* waarin men tijdens het secundair onderwijs schoolloopt langetermijneffecten heeft op de kans op werkloosheid tijdens de loopbaan. We vinden hierbij dat hoe hoger de *track*, hoe lager de kans op werkloosheid, ook wanneer we enkel vergelijkbare leerlingen in verschillende *tracks* onderzoeken. We vinden ook geen korte termijn voordeel voor het TO of BV.

Noot 1, vergelijking met SONO-rapport Laurijssen en Glorieux (2017)

Dit rapport heeft inhoudelijke overeenkomsten met het SONO-rapport "*De arbeidsmarktperspectieven van een beroepsgerichte opleiding. Een analyse van de eerste jaren van Vlaamse schoolverlaters op de arbeidsmarkt*" van Laurijssen en Glorieux (2017). Beide rapporten bekijken onder andere werkloosheid bij leerlingen uit verschillende onderwijsvormen met longitudinale gegevens en bijhorende geavanceerde analysetechnieken. Beide rapporten hebben ook specifiek aandacht voor de effecten van beroepsvoorbereidend onderwijs, door Laurijssen en Glorieux (2017) beroepsgerichte opleidingen genoemd. Voor het opstellen van onderzoeksvragen en hypotheses werd ook vaak dezelfde literatuur gebruikt. Laurijssen en Glorieux (2017) stellen ook vast dat algemene opleidingen voordelig zijn voor de werkzaamheidsgraad op lange termijn, een bevinding die bevestigd wordt in deze studie. Samengevat, zowel naar uitgangspunten als conclusies hebben dit rapport en het rapport van Laurijssen en Glorieux (2017) duidelijke overeenkomsten.

Er is echter één in het oog springend verschil. Laurijssen en Glorieux (2017) stellen vast dat beroepsgericht opgeleiden een korte termijn voordeel hebben voor werkzaamheidsgraad tegenover algemeen opgeleiden, iets wat we niet vinden in dit rapport. Dit is evenwel een schijnbare tegenstelling, doordat beide rapporten iets anders analyseren en een verschillende vergelijking maken.

Het eerste verschil tussen beide rapporten is de observatieperiode die in aanmerking wordt genomen om werkzaamheid of werkloosheid te definiëren. Laurijssen en Glorieux (2017) onderzoeken de gemiddelde kans op werkloosheid bij schoolverlaters, dus nadat jongeren het onderwijs hebben verlaten. Ons rapport daarentegen bekijkt de gemiddelde kans op werkloosheid vanaf januari 1995 tot en met december 2015 voor een hele leeftijdscohorte. In de periode dat jongeren onderwijs volgen, al dan niet in hoger onderwijs, beschouwen we hen als 'actief', terwijl deze periode niet meetelt in de kansberekening van Laurijssen en Glorieux (2017). Dat jongeren die afstuderen in het ASO in deze studie een relatief hoge kans hebben op *actief* zijn, terwijl jongeren die afstuderen in het ASO in de studie van Laurijssen en Glorieux (2017) een relatief lage kans hebben op *tewerkstelling* moet bijvoorbeeld in dit licht bekeken worden. De keuze om in dit onderzoek de gemiddelde kans op werkloosheid voor een leeftijdscohorte te meten in plaats van voor een schoolverlaterscohorte, is omdat we causale effecten van tracks op werkloosheid meten. Verschillen in leeftijden, zoals bij een schoolverlaterscohorte, zouden het meten van deze causale effecten verhinderen.

Zo komen we aan het tweede relevante verschil tussen beide rapporten: de groepen leerlingen die vergeleken worden. Waar in deze paper een vergelijking wordt gemaakt tussen tracks in het secundair onderwijs, onderscheiden Laurijssen en Glorieux (2017) beroepsgerichte en algemeen vormende opleidingen in het secundair onderwijs *en in het hoger onderwijs*. Wanneer Laurijssen en Glorieux (2017) dus spreken over verschillen tusen opleidingen in het secundair onderwijs, gaat het over leerlingen die geen diploma hoger onderwijs hebben behaald. Het rapport van Laurijssen en Glorieux (2017) toont daarbij dat gediplomeerden van het hoger onderwijs een veel hogere werkzaamheid hebben dan de algemeen opgeleiden van het secundair onderwijs. In onze analyses daarentegen kunnen leerlingen uit de verschillende tracks een hogere opleiding hebben gevolgd. Wanneer het volgen van een bepaalde track de kans verhoogt om een diploma hoger onderwijs te behalen, en dit op zijn beurt de werkzaamheidsgraad verhoogt, dan beschouwen wij dit immers ook als onderdeel van het track effect op werkloosheid. Het effect van een track op werkloosheid in dit rapport kan je dan ook als een "gecumuleerde effect" beschouwen. De bevinding van Laurijssen en Glorieux (2017) dat beroepsgericht opgeleide schoolverlaters zonder diploma hoger onderwijs initieel sneller werkzaam zijn dan algemeen opgeleide schoolverlaters zonder diploma hoger onderwijs is dan ook verenigbaar met dit rapport waar we een hogere kans op werkloosheid vinden voor de beroepsvoorbereidende track in vergelijking met de algemenere tracks van waaruit vaker wordt doorgestroomd naar het hoger onderwijs.

Noot 2, structuur van tracks

In dit rapport wordt soms gesproken over 'lagere' en 'hogere' richtingen. Het gevaar bestaat dat deze termen geïnterpreteerd worden alsof ze een bepaalde waardering van de onderzoekers weergeven. Het verschil tussen de onderwijsvormen in maatschappelijke waardering is immers vaak het onderwerp van felle discussies (Spruyt, 2014; Van Gasse, Van Cauteren, Vanhoof, & De

Maeyer, 2013). Waar in de publieke opinie sommige tracks soms een lager waardeoordeel krijgen dan andere, stellen ook heel wat mensen dat de onderwijsvormen gelijkwaardig zijn. Zij ergeren zich aan termen zoals 'hogere' en 'lagere' richtingen of 'hoog mikken'. Met het oog op een politiek correct taalgebruik wordt dan bevoordeeld beter gekozen voor begrippen zoals 'doorstroomgericht' of 'arbeidsmarktgericht'. Als onderzoekers willen we beamen we dat een negatief waardeoordeel niet thuishoort in een wetenschappelijke tekst.

Vaak wordt echter ook gezegd dat de hiërarchie tussen de onderwijsvormen en studierichtingen in Vlaanderen 'in de hoofden van de mensen' zit en dat het geen enkele basis heeft. Echter, de hiërarchische structuur van het Vlaams secundair onderwijs zit eigenlijk ingebakken in de structuur van het secundair onderwijs, los van een persoonlijk waardeoordeel. Er zijn twee structuurelementen die de hiërarchie bestendigen (Spruyt & Laurijssen, 2010). Ten eerste, leerlingen halen een diploma secundair onderwijs na het zesde jaar ASO, TSO of KSO, maar BSO-leerlingen moeten nog een zevende jaar volgen om een diploma te behalen. Ten tweede, ook het B-attest waarin de leerling geclausuleerd wordt voor bepaalde studierichtingen maar niet voor andere, illustreert treffend dat deze hiërarchie structureel is. Wie wilt doorstromen naar het TSO vanuit het ASO stoot bijvoorbeeld zelden op structurele barrières en wordt daar soms zelfs toe aangezet door een B-attest, terwijl wie wilt doorstromen naar het ASO vanuit het TSO wel bepaalde hordes tegenkomt. In die zin is er sprake van een strengere selectie voor de 'hogere' tracks dan voor de 'lagere'. Wanneer we in deze tekst impliceren dat er een hiërarchie is, dan verwijzen we kortom naar deze die structureel aanwezig is in het Vlaams secundair onderwijs.

Inhoud

1. Introduction

Policy makers have strong rationales for placing students into tracks tailored to student ability and interest. First, these tracks provide more homogeneous environments for students, assumed to lead to more efficient education (Hanushek & Wößmann, 2006). Second, it allows for skill specialization, meeting the many needs of the labor market (Bol & van de Werfhorst, 2013). Based on these rationales most education systems use some form of tracking.

Tracking is not without controversy however, with some considering it as a way to restrict resources and opportunities to a limited group of eligibles (Van de Werfhorst & Mijs, 2010). Studies indeed show that being allocated to a higher track is beneficial for academic performance (Hanushek & Wößmann, 2006). Accordingly, comparisons of education systems indicate that early tracking increases social inequality in academic performance (Van de Werfhorst & Mijs, 2010). However, academic performance is usually narrowly operationalized in specific academic skills (i.e. mathematics or language), whereas different tracks often focus on different skillsets, all assumed relevant for the labor market (Bol & van de Werfhorst, 2013). While it may be that higher tracks are beneficial for academic performance, this may not necessarily constitute an advantage in vocational opportunity.

Hence, this study assessed if tracks are equally effective in preparing students for the labor market. We propensity score matched comparable students across different tracks, to control for differences in student intake, and compared average unemployment using event history analysis. In the following sections, literature on tracking and their effects on vocational outcomes are discussed.

1.1. Tracking and inequality in academic performance

Tracking is usually understood as the ability-grouping of students into different educational programs called tracks (OECD, 2012; Van de Werfhorst & Mijs, 2010). An OECD-report (OECD, 2012, p.57-58) shows that most education systems track students during secondary education, but differ in implementation. The most notable difference is when tracking starts, with the earliest tracking at age 10 and the latest tracking at age 16. Education systems with early tracking are often characterized as selective education systems, whereas those with late tracking are characterized as comprehensive education systems (Van de Werfhorst & Mijs, 2010). Selective education systems are also associated with other aspects, such as a higher number of tracks (e.g. Bol & van de Werfhorst, 2013) and more differentiation between those tracks (Shavit & Müller, 2000).

The prevalence of tracking across education systems raises the question why students are tracked. Two rationales are typically given in the literature. The first rationale considers tracking as a way to create learning environments tailored to different learning groups. For these tracks create more homogeneous student groups, with the opportunity of focusing curricula and teachers on specific learning needs (e.g. Hanushek & Wößmann, 2006). This differentiation of students also allows for skill specialization, which is valued by the labor market (Bol & van de Werfhorst, 2013; Van de Werfhorst & Mijs, 2010). The second rationale states that education provides a form of social closure wherein people from different social backgrounds are separated. Consequently, differentiation of students into tracks institutionalizes social distance between groups (for a discussion see Bol & van de Werfhorst, 2013, pp. 287-289).

Many studies have investigated the effects of selective tracking on inequality and efficiency either by comparing education systems or by gauging the effects of educational reform (Van de Werfhorst & Mijs, 2010). When comparing education systems, studies have focused on differences in test score dispersions (Brunello & Checchi, 2007; Huang, van den Brink, & Groot, 2009) or the relationship between socioeconomic status and academic performance (Brunello & Checchi, 2007; Hanushek & Wößmann, 2006; Lavrijsen & Nicaise, 2015; Schütz, Ursprung, & Wößmann, 2008; Wößmann, 2008). Most studies found that education systems with selective tracking have more inequality than comprehensive education systems, with no clear relation between tracking and efficiency. When gauging the effects of educational reform, studies have generally shown that a reform towards a more comprehensive education system benefits disadvantaged students (Hall, 2012; Jakubowski, Patrinos, Porta, & Wiśniewski, 2016; Kerr, Pekkarinen, & Uusitalo, 2013; Malamud & Pop-Eleches, 2011; Piopiunik, 2014). Hence, there is an association between selective tracking and inequality in academic performance.

Boudon's (1974) theory on primary and secondary effects is typically used to explain the causal mechanisms within education systems responsible for inequality due to tracking (Van de Werfhorst & Mijs, 2010). The primary effect entails that higher performing students are more often allocated to higher tracks. The secondary effect entails that, even when controlling for academic performance, higher socioeconomic status students are more often allocated to higher tracks. There is broad empirical evidence of the existence of both primary and secondary effects (e.g. Jackson, Erikson, Goldthorpe, & Yaish, 2007). Hence, different students are educated in different learning environments concerning values and norms (Barth, Dunlap, Dane, Lochman, & Wells, 2004), teacher beliefs about their classrooms (Hallam & Ireson, 2003), teachers' (pedagogical) content knowledge (Baumert et al., 2010) and which skills are emphasized (Retelsdorf, Butler, Streblow, & Schiefele, 2010). Thus, both inequality in track allocation and differences in educational opportunities across tracks are key in explaining inequality in academic performance due to tracking.

1.2. Tracking and vocational outcomes

Assessing how tracking exacerbates social inequality by solely focusing on specific academic skills (i.e. mathematics or language) may however be an erroneous approach. For a primary goal of tracking is the development of different skillsets in different tracks (Ryan, 2003), addressing the various needs of the labor market (Bol & van de Werfhorst, 2013). In this view, not only academically focused skillsets are relevant for vocational opportunities, but other possible skillsets as well. Hence, the relation between tracking and vocational outcomes should be investigated as well.

It has been argued that vocationally focused tracks may even offer clearer pathways to employment. Accordingly, Bol and van de Werfhorst (2013) distinguished between education systems first by the percentage of students enrolled in vocationally focused tracks and second if skill development in vocational tracks is directly relevant for labor market entry. They found that it is the development of labor market skills which makes vocational tracks reduce youth unemployment, corroborating prior findings (Breen, 2005). Hence, vocational tracks may benefit employment.

Several studies have compared students in general tracks and vocational tracks to assess the effects of vocational tracks on vocational outcomes. Korpi, De Graaf, Hendrickx, and Layte (2003) showed that in Great Britain, the Netherlands and Sweden vocational tracks are beneficial for the transition to the labor market. However, once being unemployed, having followed a vocational track is detrimental. Hanushek, Schwerdt, Woessmann, & Zhang (2017) used a difference-in-difference analysis on 11 countries and found a relative employment advantage for vocational skills, but diminishing with age. This trade-off seemed especially true for vocationally focused tracks with apprenticeship. Lavrijsen and Nicaise (2017) compared tracks across 13 countries and showed benefits for early labor market entry. Thus, overall there is evidence that vocational track allocation improves early labor market outcomes, but these benefits decline over time.

In short, only assessing how tracking impacts inequality in specific academic outcomes may be a too narrow approach to inequality, given that vocational tracks may positively impact early vocational outcomes. However, the studies discussed in the former paragraphs on vocational outcomes mainly used regression models, with selection bias of student allocation in tracks likely remaining (see Miller & Chapman, 2001). A lack of attention for comparability of students across tracks also risks extrapolation. Instead, several authors have argued for quasi-experimental approaches such as (propensity score) matching (e.g. Schafer & Kang, 2008). Only Hanushek et el. (2017) used such approaches to discern the effects of tracks on vocational outcomes, but only with limited controls. Furthermore all of these studies simply looked at average (un)employment over time, not utilizing the advances made in event history analysis methodology (Allison, 2014). Hence, the combination of quasi-experimental and event history analysis methods has our preference to investigate the effects of tracks on vocational outcomes.

1.3. The current study

The goal of this study was to investigate if track allocation affects vocational outcomes within the Flemish education system. Large inequalities in academic performance in the Flemish education system (OECD, 2016, p442-p444) are at least partially attributed to its highly selective tracking system (Van Houtte & Stevens, 2015). However, its effects on vocational outcomes remain largely unknown.

In Flemish education, tracking starts at age 12, when students have to choose a secondary school (OECD, 2012, p57). Four tracks are organized: classical, modern, technical, and vocational. The classical and modern track are academically focused, with students expected to follow tertiary education after these tracks. The technical track offers pathways towards both tertiary education and the labor market. The vocational track primarily prepares for the labor market. Tracks are a class-level variable, with most schools offering multiple tracks, but not all. Each school has a specific profile in attracting different students, based on the tracks they offer. Student track choice is completely free if a student has attained a certificate of primary education. If no certificate has been obtained, the student is obliged to go to the vocational track. The tracks have a hierarchy in mean academic ability and mean socioeconomic background of students (Van Houtte, 2004). Unique to Flemish education is the flexibility in changing tracks downwards the hierarchy of tracks over time, often described as the “educational waterfall” (Boone & Van Houtte, 2013). Many students end their secondary education one track lower than the track they started in.

In this study we used a sample with a wide range of baseline covariates to account for the differential intake of students across tracks. Therefore, we propensity score matched (Schafer & Kang, 2008) comparable students across tracks, hence differential intake of students across tracks could not bias possible track effects. Furthermore we knew the monthly (un)employment states of former students across 21 years (1995-2015). Event history analysis was used to distinguish between the probability of becoming unemployed when being active (employed or in education) and the probability of becoming active when unemployed as a function of time (Goldstein, Pan, & Bynner, 2004; Steele, Goldstein, & Browne, 2004)

Our main hypothesis was that lower track allocation generally increases the probability of becoming unemployed. However, our secondary hypothesis was that track allocation to the vocational track decreases the probability of becoming unemployed early on, but that this advantage disappears over time and instead becomes detrimental. In the following sections the sample and methods are described in more detail.

2. Method

2.1. Sample

This study used data from the Flemish longitudinal LOSO-project (project Longitudinal Research in Secondary Education). This project followed a cohort of 6411 students in 57 schools who started secondary education in the school year 1990-1991. After secondary education, administrative data on unemployment status was provided by the government. 1543 students (24.07%) were excluded from the analyses due to their classes being de-tracked. 145 students (2.26%) were in arts or sports classes and were removed from the analyses as well. This resulted in a subsample of 4723 students. However, 390 students (8.26%) of the remaining subsample were also excluded from the analyses, having no information on their unemployment status. The remaining subsample consisted of 4333 students in 54 schools. 1522 students started in the classical track, with 461 students remaining in that track and 1061 students changing track. 1084 students started in the modern track, with 445 students remaining in that track and 639 students changing track. 1047 students started in the technical track, with 493 students remaining in that track and 554 students changing track. 680 students started in the vocational track and did not change track.

2.2. Treatment variable: track trajectories

While our main interest was the effects of tracks on unemployment, the high prevalence of students changing tracks downward the hierarchy causes our treatment variable to be track trajectories. We distinguished between those who followed a track completely and those who followed a track partially, leading to seven track trajectory groups. Pairwise comparisons were made between tracks that are consecutive in the hierarchy of tracks. Hence, a higher track was compared with a lower track, where in the higher track a distinction was made between the complete and the partial tracking trajectory. It was not possible to compare nonconsecutive tracks, due to the absence of comparable students. Thus, three three-way comparisons were made: the classical complete track with the classical partial track and the modern track, the modern complete track with the modern partial track and the technical track, and the technical complete track with the technical partial track and the vocational track. Note that students who were in a partial track overwhelmingly ended their secondary education one track lower in the hierarchy than their initial track. Within every three-way comparison each of the three possible pairs were compared.

2.3. Measures

2.3.1. Outcomes

The outcomes of interest were the probability of becoming unemployed when active and the probability of becoming active when unemployed, both as a function of time. Unemployment status of students was known from January 1995 until December 2015 on a monthly basis, equaling 252 months of potential unemployment status per person. Any other status is in our study was considered as having an active status, these mainly included being employed or in education, but also included official sick leave.

2.3.2. Baseline covariates

Matching literature advises using only those variables that predict both the treatment and the outcome of interest as baseline covariates. For variables which do not predict the outcome of interest do not reduce selection bias while decreasing the efficiency of the estimators (e.g. Myers et al., 2011). Table 1 gives a brief overview of the 15 covariates used during the different matching procedures, including how strong their effect is on the logit of becoming unemployed and the logit of becoming active (as described in the outcome analysis model later). These variables were obtained at the start of secondary education in 1990.

Table 1. Baseline covariates

Variable	Description	Info	B_{unemp}	B_{act}
NIQ	Factor score numerical intelligence	AT	-0.60	0.17
VIQ	Factor score verbal intelligence	AT	-0.56	0.17
SIQ	Factor score spatial intelligence	AT	-0.47	0.14
Math.	IRT-score of curriculum-relevant achievement in mathematics, based on 50 items	AT	-0.59	0.16
Dutch.	IRT-score of curriculum-relevant achievement in Dutch reading comprehension, based on 74 items	AT	-0.59	0.17
Gender	Binary variable gender student, reference category is boy	OR	0.14	0.03
Years	Binary variable whether a student is older than normally progressing students (delayed)	OR	1.34	-0.31
SES	Factor score socioeconomic status: based on occupation of the mother and the father, highest educational level of the mother and the father, monthly income of the family, and cultural capital of the family (amount of cultural activities)	PQ	-0.46	0.17
Room	Binary variable whether a student doesn't have an own room	PQ	0.77	-0.11
Analph.	Binary variable whether a student has at least one parent who is analphabetic	PQ	1.52	-0.28
Other lang.	Binary variable whether the home language is not Dutch	PQ	1.13	-0.33
Home	Variable with 4 categories describing the home the student lives in. Category 4 is reference.	PQ		
	Category 1: Apartment		0.73	-0.15
	Category 2: A terraced house		0.73	-0.17
	Category 3: Semi-detached house		0.35	-0.11
	Category 4: Detached house			
Garden	Variable with 3 categories describing the garden the student has access to at home. Category 1 is reference.	PQ		
	Category 1: Green garden			
	Category 2: Small courtyard garden		0.55	-0.17
	Category 3: No garden		0.37	-0.02
Nach.	Factor score need for achievement scale based on 34	SQ	-0.15	0.05

Variable	Description	Info	B_{unemp}	B_{act}
	items. Dutch diagnostic scale measuring student proneness for seeking challenges and independence			

Note: Info. = Information source; B_{unemp} = effect on logit of unemployment probability per month ; B_{act} = effect on logit of probability of becoming active per month; AT = Achievement Test; OR = Official Records; PQ = Parent Questionnaire; SQ = Student Questionnaire

2.4. Matching procedure

The goal of matching in this study was to construct matched datasets of students across track trajectories with equal confounder distributions (Schafer & Kang, 2008). Hence, any possible effect of a track trajectory on unemployment is due to the track allocation itself and none of the confounders (Winship & Morgan, 2007). Therefore, comparable respondents need to be found in the different treatment conditions (Holland, 1986). In our study this means finding comparable students on the observed baseline covariates in different track trajectories to construct matched datasets and asses the effect of track trajectories.

A prerequisite for matching is a sufficient overlap in the distribution of the baseline covariates between treatments, otherwise matching cannot be successful (Stuart, 2010). Hence, as mentioned before it was only possible to match students across track trajectories which are consecutive in the hierarchy of tracks. Furthermore, assessing the overlap indicates for which (sub)population conclusions can be drawn from the analyses (Stuart, 2010). The literature primarily focuses on average treatment effect (ATE) and the average treatment effect of the treated (ATT). In our study the ATE is the average effect of being in a higher track trajectory for all students in the three track trajectories of a three-way comparison. Accordingly, the ATT of a track trajectory comparison is the average effect of being in a higher track trajectory for all students of one specific track trajectory, resulting in three possible ATT's in a three-way comparison. Examining the overlap in the distributions of the baseline covariates between tracking trajectories led us to estimate the ATT of the higher partial track in each three-way comparison.

We used propensity score matching (PSM) to attain matched datasets. A student's propensity score in this study describes a student's probability of being allocated to the partial higher track trajectory. The theoretical foundation of PSM is that conditional on these propensities, the allocation of students is random (Imbens & Rubin, 2015). We estimated propensity scores using generalized boosted regression models (GBM's; McCaffrey et al., 2013). A GBM consists of an iterative process using multiple regression trees for capturing interacting and nonlinear relationships between track trajectory allocation and the baseline covariates without overfitting. Using the resulting propensity scores the higher complete track trajectory and lower track trajectory students are then given inverse probability weights to make these groups resemble the higher partial track trajectory. For PSM we used the package twang 1.5 in R 3.3.4.

After the matching procedures, balance in the matched datasets was assessed through standardized mean differences of covariates (SMD's, SD of lower track as denominator) between tracks after matching. Absolute mean SMD's should be no higher than 0.05 while absolute SMD's of specific covariates should not exceed 0.30 as a rule of thumb (Caliendo & Kopeinig, 2008).

2.5. Outcome analysis with discrete-time event history analysis

To model the probability of either becoming unemployed or becoming active we used discrete-time event history analysis (Allison, 2014). Discrete-time event history analysis is a flexible methodology for modeling the occurrence of an event and its timing when time is divided in intervals. We used an extended model predicting two events: becoming unemployed when active and becoming active when unemployed. Hence, the model describes transitions between two states (Goldstein et al., 2004; Steele et al., 2004), being active and being unemployed. The model used in this study can be formally described with the following equations:

$$logit\left[h_{ij}^{u}(t)\right] = \alpha_0^u + \alpha_1^u t + \alpha_2^u t^2 + \alpha_3^u t^3 + \beta^u x_j^u + u_j^u \qquad (1)$$

$$logit\left[h_{ij}^{e}(t)\right] = \alpha_0^e + \alpha_1^e t + \alpha_2^e t^2 + \alpha_3^e t^3 + \beta^e x_j^e + u_j^e \qquad (2)$$

$h_{ij}^u(t)$ was the predicted mean odds of becoming unemployed for person j in month i when in the active state. t describes time where a value of 1 equaled a year and a value of 1/12th equaled a month. α_0^u, α_1^u, α_2^u and α_3^u were respectively the intercept (ICu), linear slope (LSLu), quadratic slope (QSLu) and cubic slope (CSLu) parameters, describing the functional form of the logit of the monthly odds of becoming unemployed. x_j^u is the vector of covariates serving as the controls in this equation with corresponding vector of parameters β^u, predicting the logit of the monthly odds of becoming unemployed. u_j^u was the person-specific random effect describing a student's unique frailty to becoming unemployed. For $h_{ij}^e(t)$,the predicted mean odds of becoming active for person j in month i when in the unemployed state, an equivalent model was used. Furthermore, u_j^u and u_j^e were allowed to covary freely.

Several functional forms describing the logit odds of becoming unemployed or becoming active over time were compared (constant, linear, quadratic, cubic, quartic and grouped time-intervals). Based on relative fit indices (BIC and AIC) and stability of the estimation procedure, the cubic function came out most favorably. The person-specific random effects account for unobserved heterogeneity at the student-level, in event history analysis literature this is often referred to as a frailty model (Steele et al., 2004). The covariance between both random effects describes the relation between individual proneness to both events. Incorporating a vector of covariates used during propensity score estimation as covariates in this model was done to remove any potential bias that may remain after matching (i.e. double robustness; Schafer & Kang, 2008). For parameter estimation we used maximum likelihood estimation with robust standard errors, specified in Mplus 8.

To assess the effect of tracks, we used two approaches. A first approach was to let the ICu and ICe parameters vary freely between track trajectories, while holding the LSLu, QSLu, CSLu, LSLe, QSLe and CSLe parameters equal across track trajectories. This approach assumes that track effects are stable over time. The second approach was to let the ICu, LSLu, QSLu, CSLu, ICe, LSLe, QSLe and CSLe parameters vary freely between tracks. Hence the differences between track may differ over time. We estimated differences between tracks at the start of 1995 (T0), the start of 2000 (T1), the start of 2005 (T2), the start of 2010 (T3) , the start of 2015 (T4). Significance was tested by assessing whether the estimated difference between tracks differed significantly from zero (at each time point), using one degree of freedom Wald tests (Kuhn, Weston, Wing, & Forester, 2016).

2.6. Missing data

On average 6.54% of the baseline covariate data was missing. We used multiple imputation by chained equations to attain unbiased and efficient estimates for missing values (Schafer & Graham, 2002). Due to schools as clusters in our data, the multilevel pan-approach was used during imputation (Lüdtke, Robitzsch, & Grund, 2017). All 15 baseline covariates were included in the imputation model. Examining the autocorrelation functions and trace plots, convergence was reached after 20 iterations. We estimated ten imputed datasets, combining their results as described by Rubin (1987). The relative efficiencies attained for the differences between tracks ranged from 97.85% to 99.90% with an average of 99.41%. Estimation was done using mice 2.30 and pan 1.4 in R 3.3.4.

Out of the 1091916 unemployment statuses (4333 students multiplied by 252 months) 46386 had an unknown status (4.25%). To obtain unbiased and efficient estimates with these censored values, full information maximum likelihood (FIML) was incorporated into the estimation of the parameters.

3. Results

3.1. Track differences before matching

Table 2 shows the mean for each baseline covariate per track with a standard deviation of 1 and a mean of 0 in the complete sample. Overall, the hierarchy in the 7 track trajectories is reflected in the differences in IQ indicators, academic performance indicators and indicators for social and economic background. Moreover, there is a general trend of the lower tracks attracting more students who speak no Dutch at home.

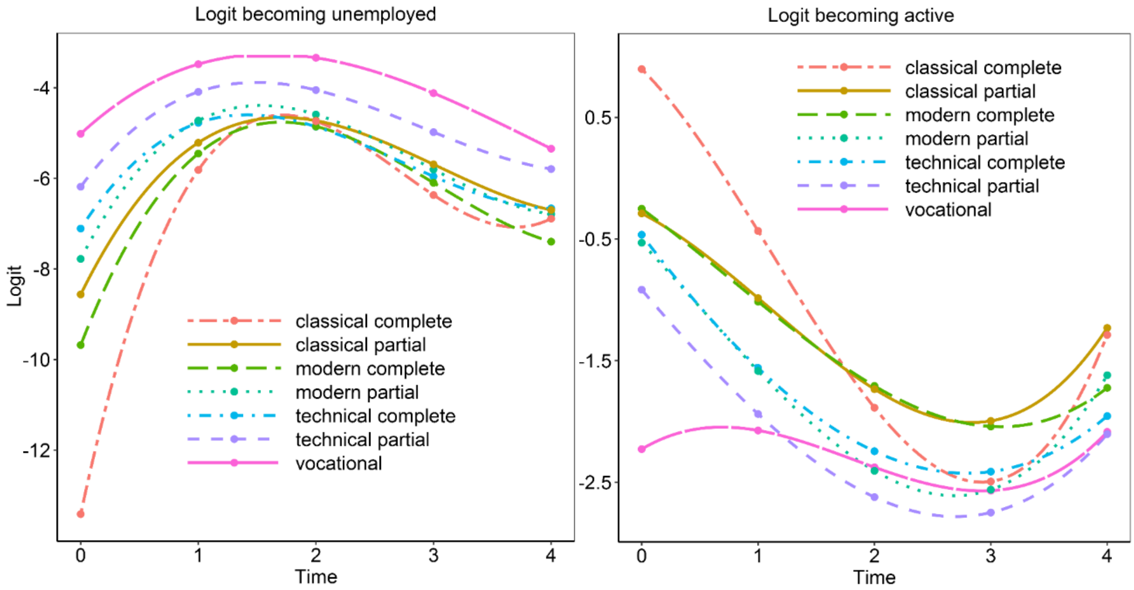
Table 2. Differences between tracks in standardized baseline covariates

Variable	Class. comp.	Class. partial	Mod. comp.	Mod. partial	Tech. comp.	Tech. partial	Vocat.
NIQ	0.86	0.47	0.43	-0.01	-0.07	-0.47	-1.21
VIQ	0.93	0.50	0.38	0.02	-0.10	-0.49	-1.27
SIQ	0.58	0.36	0.34	0.11	0.03	-0.41	-1.01
Math.	0.81	0.45	0.52	0.11	0.04	-0.41	-1.45
Dutch	1.00	0.46	0.54	0.04	-0.05	-0.56	-1.37
Gender	0.59	0.50	0.66	0.53	0.40	0.35	0.45
Lag	0.00	0.04	0.02	0.08	0.15	0.32	0.57
SES	0.85	0.40	0.23	-0.04	-0.27	-0.55	-0.80
Room	0.10	0.11	0.09	0.14	0.18	0.17	0.25
Analph.	0.00	0.02	0.00	0.01	0.03	0.05	0.18
Lang.	0.12	0.12	0.12	0.12	0.20	0.29	0.43
Home							
Cat. 1	0.03	0.04	0.02	0.03	0.04	0.05	0.06
Cat. 2	0.15	0.12	0.14	0.18	0.14	0.20	0.31
Cat. 3	0.19	0.18	0.22	0.26	0.27	0.26	0.27
Cat. 4	0.64	0.66	0.62	0.54	0.55	0.49	0.36
Garden							
Cat. 1	0.95	0.94	0.96	0.92	0.90	0.91	0.87
Cat. 2	0.04	0.04	0.03	0.06	0.07	0.07	0.11
Cat. 3	0.01	0.02	0.01	0.02	0.03	0.02	0.03
Nach.	0.35	0.10	0.19	-0.03	-0.06	-0.31	-0.22

Note: NIQ = Numerical IQ; VIQ = Verbal IQ; SIQ = Spatial IQ; Math. = Mathematics; Analph. = Analphabetism; Nach. = Need for achievement; Class. = Classical; Mod. = Modern; Tech. = Technical; Vocat. = Vocational; Comp. = Complete

Figure 1 describes the logit of becoming unemployed when active and the logit of becoming active when unemployed in the complete unmatched dataset. The logits generally reflect the hierarchy between the seven trajectories. This is most pronounced between T0 and T1 and becomes slightly fuzzier over time.

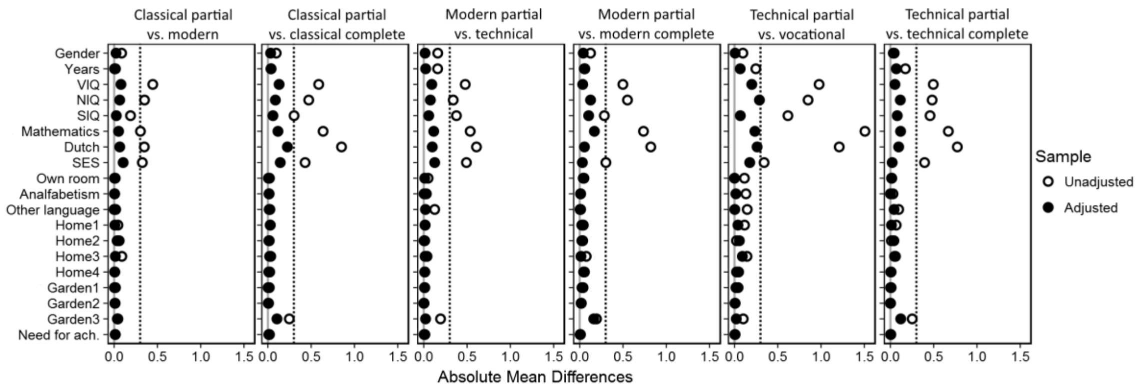
Figure 1. Event history analysis model logit becoming unemployed when active and logit becoming active when unemployed before matching.



3.2. Weighted samples after matching

Figure 2 shows all absolute SMD's for the baseline covariates relative to the reference group, the partial tracking trajectory. They show that all absolute SMD's remain under the 0.30 threshold. Furthermore the mean SMD's between tracks range from -0.03 to 0.05, not exceeding the 0.05 threshold. Thus, the intended balance in the matched datasets is reached.

Figure 2. Absolute standardized mean differences between compared groups before and after matching.



3.3. Analysis of track effects

The differences between the three track trajectories of each of the three comparisons are presented in the following paragraphs. For each comparison we describe the differences in the logits of becoming unemployed and the logits of becoming active. These differences are described for both stable differences and time-varying differences while Figure 3 and Figure 4 visualize the time-varying differences.

For the classical complete and classical partial track trajectory comparison Table 3 shows that after matching and adding covariates the time-varying differences in dunemp range from -4.23 to -0.12 while the time-varying differences in dactive range from -0.11 to 2.05. The stable differences in dunemp and dactive are -0.38 and 0.06 respectively. Concluding, for matched students the classical complete track trajectory has a lower probability of becoming unemployed, but does not differ in the probability of becoming active.

For the classical complete and modern track trajectory comparison Table 3 shows that after matching and adding covariates the differences in dunemp range from -4.13 to 0.03 while the differences in dactive range from -0.05 to 1.57. The stable differences in dunemp and dactive are -0.33 and 0.15 respectively. Concluding, for matched students the classical complete track trajectory has a lower probability of becoming unemployed, but does not differ in the probability of becoming active.

For the classical partial and modern track trajectory comparison Table 3 shows that after matching and adding covariates the time-varying differences in dunemp range from -0.02 to 0.37 while the time-varying differences in dactive range from -0.48 to 0.18. The stable differences in dunemp and dactive are 0.06 and 0.13 respectively. Concluding, for matched students the classical partial and modern track trajectory do not differ in the probabilities of becoming unemployed or becoming active.

Table 3. Predicted logits of becoming unemployed or becoming active per month comparing the classical complete, classical partial and modern tracks

Time	Difference becoming unemployed			Difference becoming active		
	partial - classical complete	modern - classical complete	modern - classical partial	partial - classical complete	modern - classical complete	modern - classical partial
$d_{T0\text{-}T4}$	-0.38*	-0.33*	0.06	0.06	0.15	0.13
d_{T0}	-4.23*	-4.13*	0.10	2.05	1.57	-0.48
d_{T1}	-0.62*	-0.64*	-0.02	0.42	0.57	0.15
d_{T2}	-0.12	-0.09	0.03	-0.11	0.07	0.18
d_{T3}	-0.70*	-0.52	0.18	-0.07	-0.05	0.03
d_{T4}	-0.34	0.03	0.37	0.01	0.14	0.13

Note. $d_{T0\text{-}T4}$ *= Average difference between tracks*; d_{T0} - d_{T4} *= Difference between tracks at given time*
* Significant at α = 0.05

For the modern complete and modern partial trajectory track comparison Table 4 shows that after matching and adding covariates the differences in dunemp range from -2.04 to -0.02 while the differences in dactive range from -0.91 to 0.54. The stable raw differences in dunemp and dactive are -0.27 and 0.32 respectively. Concluding, for matched students the modern complete track trajectory does not differ in the probability of becoming unemployed, but does have a higher probability of becoming active.

For the modern complete and technical track trajectory comparison Table 4 shows that after matching and adding covariates the time-varying differences in dunemp range from -2.50 to 0.05 while the time-varying differences in dactive range from -0.81 to 0.52. The stable differences in dunemp and dactive are -0.24 and 0.37 respectively. Concluding, for matched students the modern complete track trajectory does not differ in the probability of becoming unemployed, but does have a higher probability of becoming active.

For the modern partial and technical track trajectory comparison Table 4 shows that after matching and adding covariates the time-varying differences in dunemp range from -0.45 to 0.14 while the time-varying differences in dactive range from -0.01 to 0.42. The stable differences in dunemp and dactive are 0.03 and 0.05 respectively. Concluding, for matched students the modern partial and technical track trajectory do not differ in the probabilities of becoming unemployed or becoming active.

Table 4. Predicted logits of becoming unemployed or becoming active per month comparing the modern complete, modern partial and technical tracks

Time	Difference becoming unemployed			Difference becoming active		
	partial - modern complete	technical - modern complete	technical - modern partial	partial - modern complete	technical - modern complete	technical - modern partial
$d_{T0\text{-}T4}$	-0.27	-0.24	0.03	0.32*	0.37*	0.05
d_{T0}	-2.04*	-2.50*	-0.45	-0.91	-0.81	0.11
d_{T1}	-0.43*	-0.42*	0.01	0.37*	0.42*	0.05
d_{T2}	-0.10	0.05	0.14	0.54*	0.52*	-0.01
d_{T3}	-0.23	-0.19	0.03	0.00	0.06	0.06
d_{T4}	-0.02	-0.25	-0.23	-0.80	-0.38	0.42

Note. $d_{T0\text{-}T4}$ *= Average difference between tracks*; d_{T0} - d_{T4} *= Difference between tracks at given time*
* Significant at α = 0.05

For the technical complete and technical partial track trajectory comparison Table 5 shows that after matching and adding covariates the time-varying differences in dunemp range from -1.32 to -0.37 while the time-varying differences in dactive. range from -0.31 to 0.23. The stable differences in dunemp and dactive are -0.48 and 0.14 respectively. Concluding, for matched students the technical complete track trajectory has a lower probability of becoming unemployed, but does not differ in the probability of becoming active.

For the technical complete and vocational track trajectory comparison Table 5 shows that after matching and adding covariates the time-varying differences in dunemp range from -1.67 to -0.56 while the time-varying differences in dactive range from -0.60 to 1.18. The stable differences in dunemp and dactive are -0.73 and 0.10 respectively. Concluding, for matched students the technical complete track trajectory has a lower probability of becoming unemployed, but does not differ in the probability of becoming active.

For the technical partial and vocational track trajectory comparison Table 5 shows that after matching and adding covariates the time-varying differences in dunemp range from -0.35 to -0.18 while the time-varying differences in dactive range from -0.30 to 0.95. The stable differences in dunemp and dactive are -0.26 and -0.05 respectively. Concluding, for matched students the technical partial and vocational track trajectory do not differ in the probability of becoming unemployed and the probability of becoming active.

Table 5. Predicted logits of becoming unemployed or becoming active per month comparing the technical complete, technical partial and vocational tracks

Time	Difference becoming unemployed			Difference becoming active		
	partial - technical complete	vocational - technical complete	vocational - technical partial	partial - technical complete	vocational - technical complete	vocational - technical partial
$d_{T0\text{-}T4}$	-0.48*	-0.73	-0.26	0.14	0.10	-0.05
d_{T0}	-0.52	-1.14*	-0.62	0.23	1.18*	0.95*
d_{T1}	-0.43*	-0.64*	-0.21	0.13	0.12	0.00
d_{T2}	-0.37	-0.56*	-0.18	0.19	-0.01	-0.20
d_{T3}	-0.59	-0.90*	-0.31	0.15	0.01	-0.14
d_{T4}	-1.32*	-1.67*	-0.35	-0.31	-0.60	-0.30

Note. $d_{T0\text{-}T4}$ *= Average difference between tracks*; d_{T0} - d_{T4} *= Difference between tracks at given time*
* Significant at α = 0.05

Figure 3. Estimated logits becoming unemployed when active after matching.

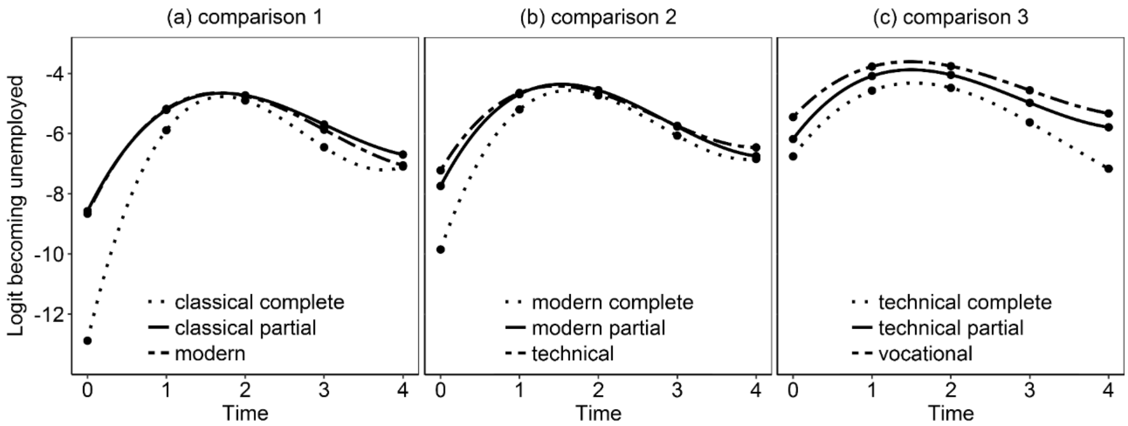
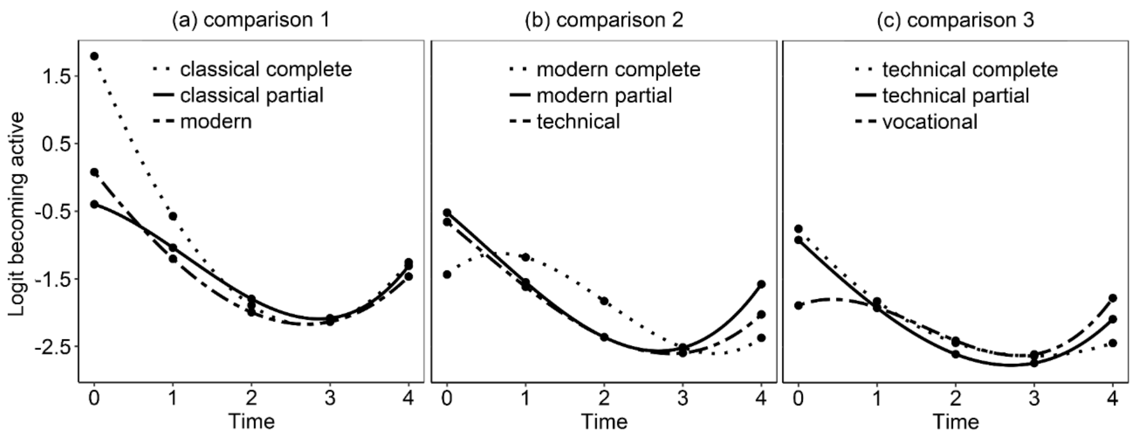


Figure 4. Estimated logits becoming active when unemployed after matching.



4. Discussion

This study investigated whether tracks in secondary education have long-term effects on the probabilities of becoming unemployed when active and becoming active when unemployed. These probabilities were compared between four tracks. Students were matched across every pair of tracks which are hierarchically consecutive, where in the hierarchically higher track a distinction was made between the complete and the partial tracking trajectory.

Supporting our hypothesis, students who followed a higher track completely had a lower probability of becoming unemployed for all comparisons. Only for one out of three comparisons students who followed a higher track completely had a higher probability of becoming active, with the reverse never occurring. Students who followed a higher track but changed to a lower track (partial trajectory) generally had the same probabilities as the other lower track students.

Assessing the differences in effect sizes reveals a more nuanced picture. The largest average difference between a higher and lower track is between the technical complete track trajectory versus the vocational track with a logit value of -0.73. This indicates that having been in a higher track leads to 0.48 times the monthly odds of becoming unemployed. In comparison, the smallest average difference is between the modern complete track trajectory and the technical track with a value of -0.24. In this case having been in a higher track leads to having 0.79 times the monthly odds of becoming unemployed. While both are tangible differences, the former difference seems much more substantial. Furthermore, the effect sizes of the differences between tracks seem more pronounced at the start and at the end of the 21-year timespan.

Concerning the probability of becoming active when unemployed, only for the modern complete track trajectory and technical track comparison a tangible difference is found. The logit value 0.37 corresponds to the higher track student having 1.44 times the monthly odds of becoming active when unemployed. Interestingly, this shows that being unemployed may not arise solely from a higher probability of becoming unemployed, but also due to a lower probability of becoming active again. Only comparing means of unemployment would not have revealed this underlying process giving rise to the unemployment data, indicating the added value of event history analysis (Allison, 2014).

Prior quasi-experimental studies on track effects have mainly focused on academic performance, finding that tracks exacerbate differences between students (Guill, Lüdtke, & Köller, 2016; Korthals & Dronkers, 2016; Retelsdorf, Becker, Köller, & Möller, 2012). This study extends those findings, showing long-term effects of tracks on unemployment. It is enticing to think that early track effects on academic performance could have caused these differences in unemployment, but we doubt this is the sole reason. Rather, we think that due to differences across tracks in skillset focus (Kunter & Baumert, 2006; Retelsdorf et al., 2010) and environments (Barth et al., 2004; Hallam & Ireson, 2003), different tracks also orient students towards different careers.

However, our result are partially at odds with prior studies on vocational tracks and unemployment. While our study agrees with most results that having been in a lower track or vocationally oriented track relates to long-term unemployment, our results do not support the early benefits as in other studies (Hanushek et al., 2017; Korpi et al., 2003; Lavrijsen & Nicaise, 2017). However, several authors already suggested that a vocational track is only effective for labor market entry if such track is strongly focused on labor market skills and apprenticeship (Bol & van de Werfhorst, 2013; Breen, 2005; Hanushek et al., 2017). Given that the vocational track in Flanders had no such focus during the nineties, our findings could fit that narrative. Furthermore, our operationalization of unemployment differs from other studies, mainly that being in education is for us not equal to unemployment. Hence, at the start of our study there is almost no unemployment. However, in other studies students start in an unemployed state before finding employment (Hanushek et al., 2017; Korpi et al., 2003; Lavrijsen & Nicaise, 2017). Hence, our finding that vocational track students are more unemployed early on and the contrary finding from other studies that vocational track students are less unemployed early on may be explained by vocational track students simply leaving education earlier.

5. Limitations and strengths

This study makes causal inferences, based on the assumption that the effects of all confounders between track allocation and the probabilities of becoming unemployed or becoming active have been accounted for through matching (ignorable treatment assumption; Rosenbaum & Rubin, 1983). Thus, a track effect can only be attributed to a track if all relevant selection bias due to confounders is removed (Steiner & Cook, 2014), an assumption which cannot be directly tested. We think though that the wide range of performance indicators and indicators of socioeconomic background to control for selection bias should have removed most selection bias. We only used need for achievement as a psychosocial control variable, for other available psychosocial variables had no tangible predictive power. Furthermore, controlling for the covariates in the outcome analysis model in our matched dataset had a negligible impact on our results, indicating the success of the matching procedure for the used variables. Hence, we think the ignorable treatment assumption is defensible.

Any estimate deriving from a matched dataset is also limited in inference to the area of common support for which enough statistical power exists in both groups (Stuart, 2010). We chose the ATT for the students of the partial tracking trajectory of the higher track in each three-way comparison. For this population of students was adequately represented in each group of a three-way comparison. Furthermore, we think that focusing the analyses on the students who are on the proverbial "fence" for track allocation are the most interesting population to investigate track effects. Our limited area of common support is also not unique to his study, for former studies on track effects on academic performance show even smaller overlap (e.g. Becker et al., 2012; Guill et al., 2016; Retelsdorf et al., 2012).

6. Conclusion

Assessing whether tracks affect the probability of becoming unemployed after secondary education revealed that having been in a higher track reduces the probability of becoming unemployed. Starting in a higher track and then changing to a lower track does not impact the probability of becoming unemployed compared to students who started in a lower track. The vocational track does not offer any benefit concerning early labor market entry. These results progress research on the effects of tracks, showing long-term tracks effects on becoming unemployed.

Bibliografie

Allison, P. D. (2014). *Event history and survival analysis Second Edition*. London, United Kingdom: Sage Publications.

Barth, J. M., Dunlap, S. T., Dane, H., Lochman, J. E., & Wells, K. C. (2004). Classroom environment influences on aggression, peer relations, and academic focus. *Journal of School Psychology*, *42*(2), 115–133.

Baumert, J., Kunter, M., Blum, W., Brunner, M., Voss, T., Jordan, A., … Tsai, Y.-M. (2010). Teachers' Mathematical Knowledge, Cognitive Activation in the Classroom, and Student Progress. *American Educational Research Journal*, *47*(1), 133–180.

Becker, M., Lüdtke, O., Trautwein, U., Köller, O., & Baumert, J. (2012). The differential effects of school tracking on psychometric intelligence: Do academic-track schools make students smarter? *Journal of Educational Psychology*, *104*(3), 682–699.

Bol, T., & van de Werfhorst, H. G. (2013). Educational Systems and the Trade-Off between Labor Market Allocation and Equality of Educational Opportunity. *Comparative Education Review*, *57*(2), 285–308.

Boone, S., & Van Houtte, M. (2013). Why are teacher recommendations at the transition from primary to secondary education socially biased? A mixed-methods research. *British Journal of Sociology of Education*, *34*(1), 20–38.

Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in western society.* New York City, New York: Wiley.

Breen, R. (2005). Explaining cross-national variation in youth unemployment: Market and institutional factors. *European Sociological Review*, 125–134.

Brunello, G., & Checchi, D. (2007). School Tracking and Equality of Opportunity. *Economic Policy*, (10), 781–861.

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72.

Goldstein, H., Pan, H., & Bynner, J. (2004). A flexible procedure for analyzing longitudinal event histories using a multilevel model. *Understanding Statistics*, *3*(2), 85–99.

Guill, K., Lüdtke, O., & Köller, O. (2016). Academic tracking is related to gains in students' intelligence over four years: Evidence from a propensity score matching study. *Learning and Instruction*, *47*, 43–52.

Hall, C. (2012). The effects of reducing tracking in upper secondary school evidence from a large-scale pilot scheme. *Journal of Human Resources*, *47*(1), 237–269.

Hallam, S., & Ireson, J. (2003). Secondary school teachers' attitudes towards and beliefs about ability grouping. *British Journal of Educational Psychology*, *73*(3), 343–356.

Hanushek, E. A., Schwerdt, G., Woessmann, L., & Zhang, L. (2017). General Education, Vocational Education, and Labor-Market Outcomes over the Lifecycle. *Journal of Human Resources*, *52*(1), 48–87.

Hanushek, E. A., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? differences-in-differences evidence across countries. *The Economic Journal*, *116*(510), C63–C76.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, *81*(396), 945–960.

Huang, J., van den Brink, H. M., & Groot, W. (2009). A meta-analysis of the effect of education on social capital. *Economics of Education Review*, *28*(4), 454–464.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge, United Kingdom: University Press.

Jackson, M., Erikson, R., Goldthorpe, H., & Yaish, M. (2007). Primary and secondary effects in class differentials in educational attaintment: the transition to A-level courses in England and Wales. *Acta Sociologica*, *50*(3), 211–229.

Jakubowski, M., Patrinos, H. A., Porta, E. E., & Wiśniewski, J. (2016). The effects of delaying tracking in secondary school: evidence from the 1999 education reform in Poland. *Education Economics*, *24*(6), 557–572.

Kerr, S. P., Pekkarinen, T., & Uusitalo, R. (2013). School Tracking and Development of Cognitive Skills. *Journal of Labor Economics*, *31*(3), 577–602.

Korpi, T., De Graaf, P., Hendrickx, J., & Layte, R. (2003). Vocational training and career employment precariousness in Great Britain, the Netherlands and Sweden. *Acta Sociologica*, *46*(1), 17–30.

Korthals, R. A., & Dronkers, J. (2016). Selection on performance and tracking. *Applied Economics*, *48*(30), 2836–2851.

Kunter, M., & Baumert, J. (2006). Who is the expert? Construct and criteria validity of student and teacher ratings of instruction. *Learning Environments Research*, *9*(3), 231–251.

Laurijssen, I., & Glorieux, I. (2017). *De arbeidsmarktperspectieven van een beroepsgerichte opleiding. Een analyse van de eerste jaren van Vlaamse schoolverlaters op de arbeidsmarkt.* Gent.

Lavrijsen, J., & Nicaise, I. (2015). New empirical evidence on the effect of educational tracking on social inequalities in reading achievement. *European Educational Research Journal*, *14*(3), 206–

221.

Lavrijsen, J., & Nicaise, I. (2017). Returns on vocational education over the life cycle: Between immediate labour market preparation and lifelong employability. *International Review of Education*, *63*(2), 257–280.

Lüdtke, O., Robitzsch, A., & Grund, S. (2017). Multiple Imputation of Missing Data in Multilevel Designs: A Comparison of Different Strategies. *Psychological Methods*, *22*(1), 141–165.

Malamud, O., & Pop-Eleches, C. (2011). School tracking and access to higher education among disadvantaged groups. *Journal of Public Economics*, *95*(11–12), 1538–1549.

McCaffrey, D. F., Griffin, B. A., Almirall, D., Slaughter, M. E., Ramchand, R., & Burgette, L. F. (2013). A tutorial on propensity score estimation for multiple treatments using generalized boosted models. *Statistics in Medicine*, *32*(19), 3388–3414.

Miller, G. A., & Chapman, J. P. (2001). Misunderstanding analysis of covariance. *Journal of Abnormal Psychology*, *110*(1), 40–48.

Myers, J. A., Rassen, J. A., Gagne, J. J., Huybrechts, K. F., Schneeweiss, S., Rothman, K. J., … Glynn, R. J. (2011). Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology*, *174*(11), 1213–1222.

OECD. (2012). Equity and Quality in Education. Paris, France: OECD.

OECD. (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. Paris, France: OECD.

Piopiunik, M. (2014). The effects of early tracking on student performance: Evidence from a school reform in Bavaria. *Economics of Education Review*, *42*, 12–33.

Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, *82*(4), 647–671.

Retelsdorf, J., Butler, R., Streblow, L., & Schiefele, U. (2010). Teachers' goal orientations for teaching: Associations with instructional practices, interest in teaching, and burnout. *Learning and Instruction*, *20*(1), 30–46.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. New York City, New York: Wiley.

Ryan, P. (2003). Evaluating Vocationalism. *European Journal of Education*, *38*(2), 147–162.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, *7*(2), 147–177.

Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: a practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313.

Schütz, G., Ursprung, H. W., & Wößmann, L. (2008). Education Policy and Equality of Opportunity. *Kyklos*, *61*(2), 279–308.

Shavit, Y., & Müller, W. (2000). Vocational secondary education, tracking, and social stratification. In M. Hallinan (Ed.), *Handbook of the sociology of education* (pp. 437–452). New York City, New York: Plenum.

Spruyt, B. (2014). De onderwaardering van technisch en beroepsonderwijs: oorsprong en gevolgen. In I. Nicaise, B. Spruyt, M. Van Houtte, & D. Kavadias (Eds.), *Het onderwijsdebat. Waarom de hervorming van het secundair broodnodig is*. Antwerp, Belgium: EPO.

Spruyt, B., & Laurijssen, I. (2010). Keuzes maken verschillen. Over de rol die onderwijskeuzes spelen in de sociale reproductie van onderwijsongelijkheid in het licht van de hervormingsvoorstellen voor het secundair onderwijs. *Tijdschrift Voor Onderwijsrecht En Onderwijsbeleid*, *11*(3), 186–195.

Steele, F., Goldstein, H., & Browne, W. (2004). A general multilevel multistate competing risks model for event history data, with an application to a study of contraceptive use dynamics. *Statistical Modelling*, *4*(2), 145–159.

Steiner, P. M., & Cook, T. D. (2014). Matching and propensity scores. In T. D. Litle (Ed.), *The oxford handbook of quantitative methods* (pp. 237–259). Oxford, United Kingdom: Oxford University Press.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, *25*(1), 1–21.

Van de Werfhorst, H. G., & Mijs, J. J. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, *36*, 407–428.

Van Gasse, R., Van Cauteren, C., Vanhoof, J., & De Maeyer, S. (2013). Maatschappelijk aanzien van onderwijsvormen. Feit of mythe? *Tijdschrift Voor Onderwijsrecht En Onderwijsbeleid*, 42–52.

Van Houtte, M. (2004). Tracking effects on school achievement: A quantitative explanation in terms of the academic culture of school staff. *American Journal of Education*, *110*(4), 354–388.

Van Houtte, M., & Stevens, P. A. (2015). Tracking and sense of futility: the impact of between-school tracking versus within-school tracking in secondary education in Flanders (Belgium). *British Educational Research Journal*, *41*(5), 782–800.

Winship, C., & Morgan, S. (2007). *Counterfactuals and causal inference*. Cambridge, United Kingdom: Cambridge University Press.

Wößmann, L. (2008). Efficiency and equity of European education and training policies. *International Tax and Public Finance*, *15*(2), 199–230.