



university of
 groningen

faculty of behavioural
and social sciences

In een Goede Flow naar het VO:

Een Onderzoek naar de Overeenstemming in Opvattingen tussen Leerling/Ouder/Leerkracht met
Betrekking tot de Vaardighedencheck

Student: L. van Drooge (S4042972)

Beoordelaar: dr. N. Frans

2^e beoordelaar: dr. A. ten Brug

Rijksuniversiteit Groningen

Faculteit der Gedrags- en Maatschappijwetenschappen

Masteropleiding Orthopedagogiek

Mei 2025

Aantal woorden: 7981

Samenvatting

In een goede flow naar het VO

Probleem- & vraagstelling: Hoewel de PO-VO overgang als cruciaal wordt beschouwd, verloopt deze niet altijd soepel. Er bestaan uiteenlopende perspectieven tussen leerkracht, ouder en leerling. Het overbruggen van de PO-VO kloof is essentieel en kan worden ondersteund door inzet van de Vaardighedencheck, die de verschillende perspectieven in kaart brengt. Dit onderzoek richt zich op het vaststellen in hoeverre de opvattingen van leerling/ouder/leerkracht van elkaar verschillen en in hoeverre ze dichterbij elkaar komen na een bespreking. **Methode:** Een correlationeel kwantitatief onderzoeksopzet in combinatie met een gemakssteekproef werd toegepast. Voor de eerste onderzoeksvraag werden data van 307 beoordelaarsparen gebruikt, voor de tweede onderzoeksvraag waren dit 33 beoordelaarsparen. De interbeoordelaarsbetrouwbaarheid is geanalyseerd met de gewogen Cohen's kappa en Krippendorff's alfa. **Resultaten:** De gemiddelde interbeoordelaarsbetrouwbaarheid tussen leerkracht en ouder is redelijk ($M = .254$), tussen ouder en leerling zwak ($M = .187$) en tussen leerkracht en leerling ook zwak ($M = .163$). De gemiddelde interbeoordelaarsbetrouwbaarheid tussen de drie beoordelaars is beperkt ($M = .247$). Daarnaast blijkt α gemiddeld .008 lager uit te vallen na herhaling; een niet-significant verschil ($p = .851$). **Conclusie:** De algemeen beperkte overeenstemming tussen leerling, ouder en leerkracht onderstreept de meerwaarde van een drieperspectievenmeting bij de Vaardighedencheck. Verder blijkt gemiddeld genomen de overeenstemming tussen leerling, ouder en leerkracht niet dichterbij elkaar te komen na herhaaldelijk gebruik, wat de noodzaak voor vervolgonderzoek benadrukt om te onderzoeken hoe meer overeenstemming bereikt kan worden binnen de Vaardighedencheck.

Sleutelwoorden: PO-VO overgang, Vaardighedencheck, drieperspectievenmeting, opvattingen, interbeoordelaarsbetrouwbaarheid

Abstract

A tool for a smooth transition to secondary school

Research problem & -questions: Although the transition from primary to secondary education is widely regarded as critical, it does not always proceed smoothly. Different perspectives between teachers, parents and students are common. Bridging this gap is essential and can be facilitated by the use of the ‘Vaardighedencheck’ (a skills-checklist). This study examines whether the perceptions of student/parent/teachers differ, and to what extent these perspectives become closer after a discussion. **Method:** A correlational quantitative design combined with a convenience sample was conducted. Data from 307 participants addressed the first research question and data from 33 participants addressed the second. Inter-rater agreement was analyzed using Weighted Cohen’s Kappa and Krippendorff’s Alpha. **Results:** On average the inter-rater agreement between teacher and parent was moderate ($M = .254$), between parent and student weak ($M = .187$), and between teacher and student also weak ($M = .163$). Inter-rater agreement among all three raters was, on average, limited ($M = .247$). Furthermore, the α was on average .008 lower after repetition, a difference that is not statistically significant ($p = .851$). **Conclusions:** The findings indicate generally limited agreement among students, parents, and teachers, highlighting the added value of including a three-perspective measurement within the ‘Vaardighedencheck’. Furthermore, repetition of the ‘Vaardighedencheck’ did, on average, not bring the three perspectives closer together which emphasizes the need for further research into how consensus might be improved.

Key words: primary-to-secondary transition, Vaardighedencheck (skills-checklist), three-perspective measurement, perceptions, inter-rater agreement

Inhoudsopgave

Introductie	5
Methode	9
Design	9
Populatie en Steekproef.....	9
Instrument en Variabelen.....	9
Procedure	10
Analyseplan	10
Resultaten	13
Missende data	13
Scores.....	13
Verdeling verschillscores.....	14
Interbeoordelaarsbetrouwbaarheid.....	15
Relatie Interbeoordelaarsbetrouwbaarheid en Scores.....	20
Vergelijking schooljaren.....	21
Conclusie/Discussie	24
Conclusie	24
Discussie	27
Literatuurlijst	32
Bijlage A.....	36
Bijlage B.....	39
Bijlage C.....	42

Introductie

Nederlandse leerlingen doorlopen doorgaans acht jaar primair onderwijs (PO) voordat zij de overgang maken naar het voortgezet onderwijs (VO). De PO-VO overgang wordt in veel studies beschouwd als cruciale verandering in het leven van een kind (Evans et al., 2018; Harris & Nowland, 2021; Naaijer et al., 2016). Echter, deze overgang verloopt niet voor iedere leerling soepel (Harris & Nowland, 2021; Naaijer et al., 2016; Van Rens, 2020). In groep 8 voeren leerkrachten adviesgesprekken met leerling en ouders waarbij het uitstroomperspectief van de leerling besproken wordt (Koren, 2024; Van Rooijen et al., 2017). Uit het onderwijsonderzoek van DUO (2020) blijkt dat circa 11.000 ouders van mening zijn dat het uitstroomperspectief te laag is, terwijl leerkrachten in het voortgezet onderwijs juist het aantal te hooggeplaatste brugklassers zien stijgen. Naast de grote invloed die didactische vaardigheden hebben op het schooladvies, spelen sociaal-emotionele en executieve vaardigheden ook een rol (Koren, 2024). Door tijdig inzichtelijk te maken in hoeverre een leerling benodigde vaardigheden beheerst, kunnen tegenstrijdige opvattingen tussen leerling, ouder en leerkracht bespreekbaar worden gemaakt. Inzicht kan worden gerealiseerd door de mening van zowel leerlingen als ouders eerder in het overgangsproces mee te nemen, zodat mogelijke onenigheid over het advies tijdig kan worden opgehelderd en er open communicatie plaatsvindt tussen de leerling, ouder en school. De actieve betrokkenheid van leerlingen en ouders bij de schoolgang vergemakkelijkt het adviesgesprek over de PO-VO overgang (Van Rens, 2020).

Steeds meer leerlingen lijken over onvoldoende sociaal-emotionele vaardigheden, motivatie en concentratie te beschikken om goed het voortgezet onderwijs te beginnen (Durlak et al., 2011) en ervaren aansluitproblematiek (Korpershoek et al., 2016). Deze aansluitproblematiek wordt ook wel aangeduid als de ‘PO-VO kloof’ (Van Gennip & Klinkenberg, 2023). Binnen het voortgezet onderwijs bestaan namelijk andere sociale verwachtingen dan leerlingen gewend zijn vanuit het primair onderwijs (Besi & Sakellariou, 2019; Harris & Nowland, 2021; Korpershoek et al., 2016). Naast nieuwe sociale verwachtingen worden ook andere studievoordigheden van leerlingen verwacht binnen het voortgezet onderwijs. In het bijzonder executieve vaardigheden zoals het maken van uitgebreide opdrachten, plannen, zelfstandig studeren en het nemen van verantwoordelijkheid zijn belangrijk (Jindal-Snape et al., 2019; Koren, 2024; Van Rens et al., 2017).

Korpershoek et al. (2016) stellen dat de aanwezigheid van voldoende vaardigheden op het gebied van cognitie, zelfstandig werken en aanpassingsvermogen een voorwaarde is voor een soepele PO-VO overgang. Scholen dienen daarom niet alleen de cognitieve vaardigheden van leerlingen te stimuleren, maar ook de sociaal-emotionele vaardigheden (Durlak et al., 2011). Verder benadrukken meerdere onderzoeken dat omgevingsaspecten ook van invloed zijn op een soepele PO-VO overgang. Vooral de interpersoonlijke relaties, zoals de relatie tussen leerling en ouders/medeleerlingen, en tussen ouders en school spelen hierbij een rol (Besi & Sakellariou, 2019; Jindal-Snape et al., 2019; Van Rens et al., 2017). Daarnaast is een adequate relatie tussen leerling en leerkracht ook belangrijk en noemen Van Rens et al. (2017) dat leerkrachten de PO-VO overgang kunnen vergemakkelijken door leerlingen aan te moedigen zelf actief deel te nemen in beslissingen die genomen worden omtrent hun overgang. Van Rens et al. (2017) benadrukken dat een succesvolle overgang naar het voortgezet onderwijs bijdraagt aan de optimale ontwikkeling van leerlingen. Bovendien draagt het bij aan het positief welzijn van leerlingen en heeft het een positieve invloed op toekomstige leerprestaties (Jindal-Snape et al., 2019; Korpershoek et al., 2016; Van Rens, 2020).

Ondanks het belang van een succesvolle overgang, de bekendheid van de mogelijke problemen tijdens dit proces en de complexiteit ervan, blijft onderzoek omtrent het onderwerp beperkt (Smith et al., 2008, Van Rens, 2020). Er is weinig bekend over de voorbereidingen in het primair onderwijs en over eventuele aansluitingsproblemen in het voortgezet onderwijs (Driessen et al., 2005; Korpershoek et al., 2016). Er is voornamelijk weinig bekend over de meerwaarde van de rol van leerlingen tijdens het proces (Van Rens, 2020). En hoewel bekend is dat sociaal-emotionele en executieve vaardigheden belangrijk zijn bij een soepele PO-VO overgang, is het onduidelijk in hoeverre de opvattingen van leerling, ouder en leerkracht omtrent deze vaardigheden overeenkomen. Volgens Van Gennip en Klinkenberg (2023) is het overbruggen van de PO-VO kloof van groot belang. Om dit te realiseren, hebben zij het instrument ‘de Vaardighedencheck’ ontwikkeld. De Vaardighedencheck meet middels drie vragenlijsten in hoeverre een leerling de benodigde vaardigheden voor een succesvolle PO-VO overgang beheerst. Dit wordt gedaan vanuit het perspectief van de leerling, ouder en leerkracht. Er wordt onderscheid gemaakt tussen sociale- en communicatieve vaardigheden aan de ene kant en executieve- en schoolse vaardigheden aan de andere kant. Daarnaast wordt er gevraagd naar het welbevinden, de schoolaanwezigheid en de schoolresultaten van de leerling. Wanneer de drie beoordelaars

(leerling/leerkracht/ouder) de Vaardighedencheck hebben ingevuld, wordt een totaaloverzicht gecreëerd. Dit totaaloverzicht geeft inzicht in welke vaardigheden de leerling reeds beheerst en welke vaardigheden extra aandacht behoeven. De Vaardighedencheck heeft een preventieve inzet met als doel de onderwijskundige basisondersteuning te versterken en het handelingsrepertoire van leerkrachten te verbreden (Van Gennip & Klinkenberg, 2023). Daarnaast fungeert het als middel om planmatig en doelgericht werken op scholen vorm te geven. Volgens de makers van het instrument kan de Vaardighedencheck een middel zijn om de ontwikkeling van leerlingen te stimuleren, aangezien de ondersteuningsbehoeftes van de leerling expliciet in beeld gebracht worden (Van Gennip & Klinkenberg, 2023).

Bij het gebruik van verschillende perspectieven, bestaat de kans dat waargenomen gedrag op verschillende manieren geïnterpreteerd wordt. Dit levert vragen op over zowel de validiteit als de betrouwbaarheid van het instrument. Zo blijkt uit onderzoek van Murray et al. (2021) over een ander instrument met een meerdere perspectievenmeting, dat ouders vaak meer problemen signaleren dan leerkrachten. Een mogelijke verklaring voor deze verschillen is dat ouders en leerkrachten het kind in een andere omgeving meemaken en ook andere interactie hebben met het kind (Bergström & Baviskar, 2021). Uit een evaluatie van de Vaardighedencheck blijkt dat beoordelaars het als interessant ervaren om bij een verschil in beoordeling tijdens het driehoeksgesprek te verkennen of dit verschil voortkomt uit een uiteenlopende interpretatie van een bepaalde vaardigheid, of dat er daadwerkelijk sprake is van verschillend gedrag in diverse contexten (Van Gennip & Klinkenberg, 2023). Het doel van de drie perspectievenmeting is om de verschillende perspectieven tijdig aan het licht te brengen, zodat uiteenlopende perspectieven voorkomen worden en er een transparante samenwerking plaatsvindt (Van Gennip & Klinkenberg, 2023). Om deze reden is het relevant om te onderzoeken in hoeverre de opvattingen van leerling, ouder en leerkracht daadwerkelijk van elkaar verschillen en in hoeverre ze dichterbij elkaar komen na een bespreking van de resultaten. Inzichten in de mate van overeenstemming tussen de beoordelaars per item kunnen leiden tot aanpassingen aan de vraagstelling bij sterk uiteenlopende perspectieven of verwijderen van een item wanneer perspectieven perfect overeenkomen. Systematische verschillen in perspectieven op specifieke vaardigheden kunnen ook een aanleiding zijn voor scholen om deze verschillen al vroeg te ondervangen. De verwachting is dat open communicatie tussen ouder, leerling en school de perspectieven meer op één lijn brengt, dit onderzoek laat zien of een drieperspectievenmeting meerwaarde zal hebben. In dit onderzoek is

daarom gefocust op de verschillende perspectieven van leerling, ouder en leerkracht met betrekking tot de Vaardighedencheck. Huidig onderzoek kan gezien worden als deel van een groter valideringsonderzoek, waarbij psychometrische eigenschappen en het gebruik van de Vaardighedencheck geëvalueerd worden. Naast huidig onderzoek lopen er vier andere onderzoeken omtrent de Vaardighedencheck, ieder met een eigen focus. Het bredere doel van de onderzoeken is om de Vaardighedencheck te verfijnen. Het doel van huidig onderzoek is om de overeenstemming in opvattingen tussen leerling, ouder en leerkracht inzichtelijk te maken. Aan de hand van de bevindingen van dit onderzoek kan geëvalueerd worden over de items in de Vaardighedencheck en in hoeverre een drieperspectieven meting meerwaarde geeft.

In dit onderzoek staan twee onderzoeksvragen centraal. De eerste onderzoeksvraag is: “In hoeverre komen de, in de Vaardighedencheck gemeten, opvattingen over de vaardigheden van leerling/ouder/leerkracht overeen?”. De deelvragen hierbij zijn: (1) “Hoe verschillen de opvattingen tussen beoordelaars?” en (2) “Hoe verschillen de opvattingen tussen items?”. De tweede onderzoeksvraag is, “In hoeverre komen de, in de Vaardighedencheck gemeten, opvattingen over de vaardigheden van leerling/ouder/leerkracht dicht bij elkaar na een bespreking van de Vaardighedencheck?”.

Methode

Design

Tijdens het onderzoek werd een correlationeel kwantitatief onderzoeksopzet gevolgd. Het betrof een retrospectief onderzoek; de data werden verzameld in een natuurlijke setting en de variabelen werden zonder manipulatie geobserveerd. Correlationeel kwantitatief onderzoek meet het verband tussen verschillende variabelen en binnen dit onderzoek werd het ingezet om de overeenstemming (interbeoordelaarsbetrouwbaarheid) tussen leerling, ouder en leerkracht te onderzoeken. Het onderzoek betreft een retrospectief onderzoek omdat het uitgevoerd werd aan de hand van reeds bestaande data, die werden verzameld aan de hand van de online dataretour van de Vaardighedencheck.

Populatie en Steekproef

De doelpopulatie van het onderzoek waren leerlingen, ouders en leerkrachten van het primair onderwijs in Nederland. De toegankelijke populatie bestond uit leerlingen, ouders en leerkrachten voornamelijk uit groep 7 en 8 van het primair onderwijs in de regio Twente. Gezien de vrijwillige deelname van scholen en leerkrachten betrof het een gemakssteekproef. Van het schooljaar 2022-2023 waren er 307 participanten, en van het schooljaar 2023-2024 waren er 33 bruikbare participanten. Met één participant wordt in dit onderzoek een triade van ouder/leerling/leerkracht bedoeld. Voor het beantwoorden van de eerste onderzoeksvraag werden de data van schooljaar 2022-2023 gebruikt. Voor het beantwoorden van de tweede onderzoeksvraag was het inclusiecriterium dat er data beschikbaar was van zowel 2022-2023 als 2023-2024 ($n = 33$). Bij de tweede onderzoeksvraag werd de assumptie gemaakt dat er tussen de metingen een driehoeksgesprek is gevoerd.

Instrument en Variabelen

Het belangrijkste instrument binnen dit onderzoek was de Vaardighedencheck. In het kader van het onderzoek werden er geen aanpassingen aan het bestaande instrument gedaan. De Vaardighedencheck bestaat uit drie verschillende vragenlijsten: één voor de leerling (zie Bijlage A), één voor de leerkracht (zie Bijlage B) en één voor de ouder (zie Bijlage C). Elke vragenlijst bevat 44 stellingen over de leerling met betrekking tot de thema's "sociale en communicatieve

vaardigheden”, “executieve en schoolse vaardigheden” en “welbevinden, schoolaanwezigheid en schoolresultaten”. De stellingen zijn voor elke beoordelaar vergelijkbaar en elk item refereert naar bepaald gedrag dat een leerling al dan niet laat blijken (bijvoorbeeld, “Ik begin uit mezelf met het maken van huiswerk”) en een bijbehorend construct. De items zijn te beantwoorden op een vijfpunt-Likert schaal, variërend van nooit (1) tot altijd (5). De betrouwbaarheid en validiteit van de Vaardighedencheck zijn tot op heden nog niet empirisch vastgesteld. De nominale variabele in het onderzoek waren de participant en de beoordelaar, de ordinale variabele was de itemscore en de dichotome variabele was de aan-/afwezigheid van een tweede meting. De afhankelijke variabelen zijn de itemscores; waarbij onderscheid werd gemaakt tussen de verschillende beoordelaars (leerkracht/leerling/ouder).

Procedure

Data werden in de periode 2021 tot 2023 verzameld via de dataretour van de Vaardighedencheck. De data werden automatisch naar de ontwikkelaar van het instrument gestuurd, die het vervolgens doorstuurde naar de onderzoeker. Scholen in de omgeving Twente zijn benaderd om de Vaardighedencheck toe te passen en hebben zich hier vrijwillig voor aangemeld. De deelnemende scholen hebben een verwerkersovereenkomst getekend waarin de verwerking en het gebruik van de data voor wetenschappelijk onderzoek beschreven staat. Leerlingen, ouders en leerkrachten hebben de Vaardighedencheck ingevuld door in te loggen op de website met een unieke gebruikersnaam en wachtwoord. Er werd actief toestemming gegeven voor de verwerking van de data en het gebruik van geanonimiseerde data voor algemeen wetenschappelijk onderzoek door een schriftelijk akkoord van ouders. Daarnaast werd het onderzoek goedgekeurd door de ethische commissie van Pedagogische- en Onderwijswetenschappen van de Rijksuniversiteit van Groningen. Beoordelaars vulden de vragenlijst individueel en onafhankelijk van elkaar in, op een zelf gekozen fysieke locatie.

Analyseplan

De analyses van het onderzoek werden uitgevoerd aan de hand van het statistische programma IBM SPSS Statistics 28. Om te beginnen werd er gezocht naar structurele relaties in de missende data door een dummyvariabele (0=geen missende data, 1= missende data) aan te maken en de gemiddelde totaalscores van de groep zonder missende data en de groep met missende

data met elkaar te vergelijken. Dit werd gedaan door uitvoering van een onafhankelijke t -toets (significantieniveau $p < .05$). Voor de volgende analyses werden alleen observaties met missende data niet meegenomen (pairwise deletion), waardoor verzamelde data niet verloren gingen.

Om inzichtelijk te maken hoe de drie beoordelaars de verschillende items beoordelen, werd allereerst de gemiddelde itemscore per beoordelaar berekend. Op basis van deze gemiddelden werd zichtbaar hoe de scores zich onderling verhouden en welke beoordelaar gemiddeld genomen het meest positief beoordeelde. Om te onderzoeken in hoeverre de beoordelaars van elkaar verschillen in hun beoordelingen, werden vervolgens verschillscores berekend. Deze verschillen laten zien in welke mate de beoordelaarsparen het met elkaar eens zijn of juist van elkaar afwijken. Daarnaast werd er gekeken naar de absolute verschillscores, omdat hoge en lage itemverschillen elkaar kunnen compenseren. Door de absolute waarden van de verschillen te gebruiken, werd zichtbaar hoe groot de afwijkingen tussen beoordelaars zijn. Dit geeft een zuiverder beeld van de mate van overeenstemming tussen de beoordelaarsparen. Eerst werd beschrijvende analyse gebruikt om te kijken naar de verdeling van de (absolute) verschillscores per beoordelaarspaar. Om structurele itemverschillen vast te stellen werd vervolgens gekeken naar de verschillscores per item.

Om de mate van interbeoordelaarsbetrouwbaarheid tussen twee beoordelaars te meten, werd per beoordelaarspaar de gewogen Cohen's kappa berekend. Deze maat drukt de overeenstemming tussen beoordelaars uit in één getal, waarbij gecorrigeerd wordt voor overeenstemming als gevolg van toeval (Warrens, 2017). Gewogen Cohens's kappa werd verkozen boven de standaard Cohen's kappa gezien het ordinale karakter van de variabelen en de aanwezigheid van meer dan twee categorieën (Marasini et al., 2014; Yue et al., 2013). In dit onderzoek werd gekozen voor het toepassen van lineaire gewichten. Een voordeel van de gewogen kappa is dat grote beoordelingsverschillen zwaarder gewogen worden dan kleine beoordelingsverschillen. Zo weegt een discrepantie tussen de categorieën 'altijd' en 'nooit' zwaarder mee dan een verschil tussen 'altijd' en 'meestal'. Dit zorgt voor een meer genuanceerde beoordeling van de onderlinge overeenstemming. Door de gewogen Cohen's kappa afzonderlijk te berekenen voor elk beoordelaarspaar wordt inzichtelijk hoe de overeenstemming verschilt tussen leerling/ouder, leerling/leerkracht en ouder/leerkracht. Deze uitkomsten werden vervolgens samengevat in histogrammen. Ter aanvulling werd er gekeken naar de frequentie van leerkrachtbetrokkenheid omtrent de matige overeenstemmingen en onenigheid, aan de hand van een frequentietabel.

Om de mate van interbeoordelaarsbetrouwbaarheid tussen de drie beoordelaars te meten, werd de Krippendorff's alfa berekend per participant. De Krippendorff's alfa limiteert zich niet tot een aantal observatoren, kan worden gebruikt bij een ordinaal meetniveau en bij eventuele ontbrekende data (Hayens & Krippendorff, 2007). Deze analyse werd uitgevoerd met behulp van de KALPHA macro (Hayens & Krippendorff, 2007), in SPSS. Het doel van de gewogen Cohen's kappa en Krippendorff's alfa analyse is om antwoord te krijgen op de eerste onderzoeksvraag. De interpretaties van de waarden zijn weergegeven in Tabel 1.

Om de overeenstemming tussen alle drie beoordelaars tijdens de eerste en de tweede meting te toetsen en daarmee antwoord te krijgen op de tweede onderzoeksvraag, werd een nieuwe dataset aangemaakt. Daarin werd de daarvoor berekende algehele overeenstemming (KALPHA) gebruikt als afhankelijke variabele. Een gepaarde *t*-test (significantieniveau $p < .050$) werd ingezet om de overeenstemming te toetsen. Om meer duiding op deze hoofdvraag te krijgen werd ook gekeken of er een verschil in de hoogte van de totaalscores bestaat tussen hoge en lage overeenkomstsscores. Hiervoor werd een scatterplot gemaakt die de hoogte van Krippendorff's alfa afzet tegen de hoogte van de totaalscore. Ter aanvulling werd er wederom aan de hand van een frequentietabel gekeken naar leerkrachtbetrokkenheid omtrent hogere/lagere overeenstemmingen na een bespreking.

Tabel 1

Interpretaties van Gewogen Cohen's Kappa (Cohen, 1968) en Krippendorff's Alfa (Krippendorff, 2004)

Interpretatie	Kappa-waarde	Alfa-waarde
Zwakke overeenstemming	< .20	< .667
Redelijke overeenstemming	.20-.40	.667-.80
Matige overeenstemming	.41-.60	
Sterke overeenstemming	.61-.80	> .80
Zeer sterke overeenstemming	>.80	

Resultaten

Missende data

In Tabel 2 wordt een beschrijvend overzicht gepresenteerd van beantwoorde en missende items. Bij de analyse van missende data valt op dat er verschillen zijn in het aantal beantwoorde items door verschillende beoordelaars. Ouders vullen 25% van de items niet in, terwijl dit percentage voor zowel leerkrachten als leerlingen onder de 10% ligt. Dit hoge percentage onbeantwoorde items door ouders komt voornamelijk doordat ouders de volledige vragenlijst niet invullen en bij leerlingen en leerkrachten is het vaker het geval dat er sporadisch een item niet wordt beantwoord. Op basis van de scores wordt geen structurele relatie tussen beide groepen verwacht. Er is namelijk geen significant verschil gevonden in gemiddelde scores tussen de groepen met ($M = 3.527$) en zonder missende data ($M = 3.557$), $t(305) = 0.586$, $p = .558$.

Tabel 2

Beschrijvend overzicht van Beantwoorde en Onbeantwoorde items, uitgesplitst naar Beoordelaar

	Beantwoorde items	Onbeantwoorde items
	<i>n</i> (%)	<i>n</i> (%)
Leerling	19922 (92.0%)	1726 (8.0%)
Leerkracht	20086 (92.8%)	1562 (7.2%)
Ouder	16082 (74.3%)	5566 (25.7%)
Gemiddelde totaalscore	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)
	3.557 (.037)	3.527 (.034)

Scores

In Tabel 3 wordt de gemiddelde itemscore en mediaan per beoordelaar gepresenteerd. Ouders blijken gemiddeld genomen lagere scores te rapporteren dan zowel leerkrachten als leerlingen, wat betekent dat ouders over het algemeen negatiever zijn. De mediaan voor ouders is 3, terwijl zowel leerkrachten als leerlingen een mediaan van 4 rapporteren. Dit betekent dat meer dan 50% van de ouders het vaakst een score van 3 of lager rapporteren, terwijl leerkrachten en leerlingen meer dan 50% van de tijd een score van 4 of hoger rapporteren.

Tabel 3

Gemiddelde Itemscore en Mediaan per Beoordelaar

	<i>M (SD)</i>	<i>Mdn</i>
Leerling	3.62 (1.141)	4
Leerkracht	3.56 (1.074)	4
Ouder	3.42 (1.141)	3

Verdeling verschillscores

In Tabel 4 wordt het (absolute) gemiddelde verschil tussen de beoordelaarsparen gepresenteerd. De gemiddelde verschillscores bevestigen dat ouders negatiever rapporteren dan zowel leerlingen als leerkrachten, en dat leerlingen positiever rapporteren dan leerkrachten. Hoewel uit de verschillscore blijkt dat leerkrachten en leerlingen gemiddeld genomen niet structureel lijken af te wijken, blijkt uit de absolute verschillscore dat zij het minder met elkaar eens zijn dan leerkrachten en ouders.

Tabel 4

(Absolute) Gemiddelde Verschilscores tussen Beoordelaarsparen

	Verschilscore	Absolute verschillscore
	<i>M (SD)</i>	<i>M (SD)</i>
Leerkracht-ouder	.130 (1.150)	.819 (.299)
Leerkracht-leerling	-.075 (1.272)	.945 (.283)
Ouder-leerling	-.212 (1.279)	.958 (.288)
Aantal geldige scores	<i>n</i>	<i>n</i>
	10516	10516

De verschillscores van de meeste items hadden een mediaan van 0. Een afwijkende mediaan (*Mdn* ≠ 0) betekent dat er een structureel verschil in opvatting is bij het desbetreffende item. De verschillscores tussen leerkracht en ouder tonen dat voor enkel item 22 (uit zichzelf beginnen met huiswerk maken) de mediaan afwijkt, waarbij de leerkracht structureel positiever is dan de ouder. Bij de verschillscores tussen leerling en leerkracht wijkt enkel item 34 (oorzaken bij jezelf zoeken)

af, waarbij de leerkracht structureel minder positief is dan de leerling. Opvallend is dat voor leerling en ouder het structurele verschil in opvatting is gevonden bij zes items. Voor de items 10 (feedback geven), 29 (planning aanpassen), 30 (doorzetten bij moeilijkheid), 31 (doorzetten bij teleurstelling) en 34 (oorzaken bij jezelf zoeken) is de ouder minder positief dan de leerling. Voor item 39 (zin om naar school te gaan) is de ouder juist positiever dan de leerling. De volledige vraagstellingen per item zijn te vinden in Bijlage A, B en C.

Interbeoordelaarsbetrouwbaarheid

De proportie van exacte overeenstemming tussen de beoordelaarsparen wordt gepresenteerd in Tabel 5. Hierbij is te zien wat de proportie exacte overeenstemming per itemscore is en wat het gemiddelde is over deze proporties. Het percentage items waarbij de antwoorden exact overeenkomen, is voor twee van de drie beoordelaarsparen het hoogst bij de score 4. De gemiddelde exacte overeenstemming tussen leerkracht en ouder is het hoogst, en de gemiddelde exacte overeenstemming tussen leerkracht en leerling is het laagst.

Tabel 5

Proportie Exacte Overeenstemming tussen Alle Beoordelaars, zowel per Score als Gemiddeld.

Score	Leerkracht/ouder	Ouder/leerling	Leerkracht/leerling
1	15.80%	12.20%	8.40%
2	39.30%	36.40%	28.70%
3	34.40%	32.00%	32.30%
4	42.50%	37.30%	38.30%
5	44.70%	34.90%	37.10%
Gemiddeld	35.34%	30.56%	28.96%
Aantal geldige scores	<i>n</i>	<i>n</i>	<i>n</i>
	15135	15185	18839

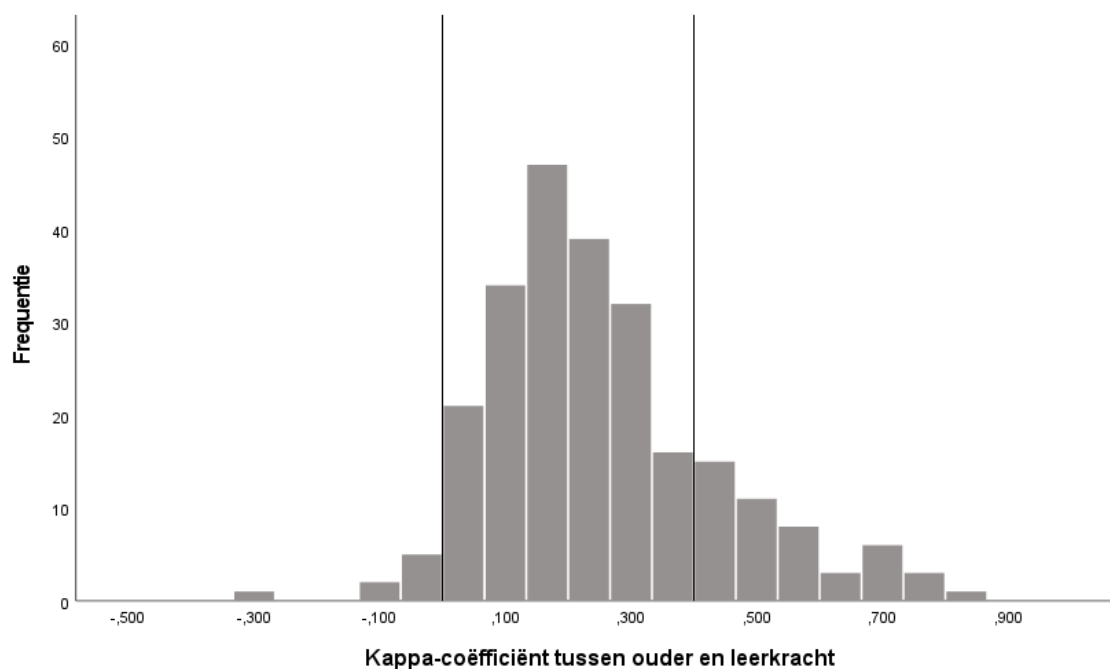
Noot. Proportie berekend op basis van het aantal keer dat beide beoordelaars een score toekennen gedeeld door het aantal toekenningen van de eerste beoordelaar in het paar.

Figuur 1, Figuur 2 en Figuur 3 presenteren de gewogen Cohen's kappa-waarde in histogrammen, met bijbehorende afkapscorelijnen. De kappa toont aan dat er gemiddeld een redelijke interbeoordelaarsbetrouwbaarheid is tussen ouder en leerkracht ($M = .254$, $SD = .181$, $n = 244$) en een zwakke interbeoordelaarsbetrouwbaarheid tussen zowel leerling en ouder ($M = .187$, $SD = .139$, $n = 240$) als tussen leerling en leerkracht ($M = .163$, $SD = .137$, $n = 289$).

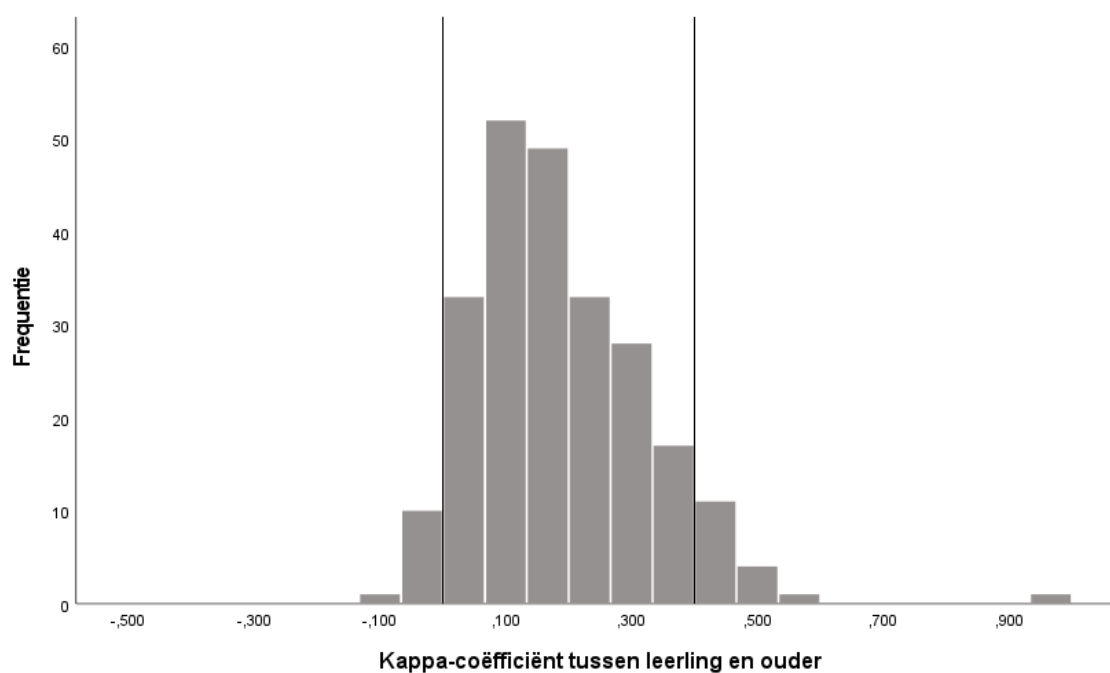
De interbeoordelaarsbetrouwbaarheid wordt in de meeste gevallen gekarakteriseerd als zwak tot redelijk. Voor enkele leerlingen is echter een systematische onenigheid gevonden ($\kappa < 0$), terwijl voor andere leerlingen de interbeoordelaarsbetrouwbaarheid matig is ($\kappa > .40$). Om te zien of er een verband bestaat tussen de mate van overeenstemming en de totaalscore van de Vaardighedencheck, wordt de range van de totaalscores van de drie beoordelaars gegeven, waarbij de minimumscore 132 is en de maximumscore 660. Bij de overeenstemming tussen ouder en leerkracht zijn er 47 gevallen waarbij de interbeoordelaarsbetrouwbaarheid matig is; hierbij liggen de totaalscores tussen de 303 en 579. In acht gevallen is er onenigheid tussen ouder en leerkracht; waarbij de totaalscores liggen tussen de 285 en 578. Bij de overeenstemming tussen leerling en ouders zijn er 17 gevallen waarbij de interbeoordelaarsbetrouwbaarheid matig is; hierbij liggen de totaalscores tussen de 289 en 540. In tien gevallen is er onenigheid tussen leerling en ouder gevonden; hierbij liggen de totaalscores tussen de 286 en 578. Bij de overeenstemming tussen leerling en leerkracht zijn er 11 gevallen waarbij de overeenstemming matig is; de totaalscores van deze 11 gevallen liggen tussen de 275 en 552. In 28 gevallen is er onenigheid tussen leerling en leerkracht; waarbij de totaalscores liggen tussen de 214 en 578.

Figuur 1

Gewogen Kappa-waarden voor de Overeenstemming tussen Ouder en Leerkracht, met Afkapscorelijn bij $\kappa < 0$ en $\kappa > .40$

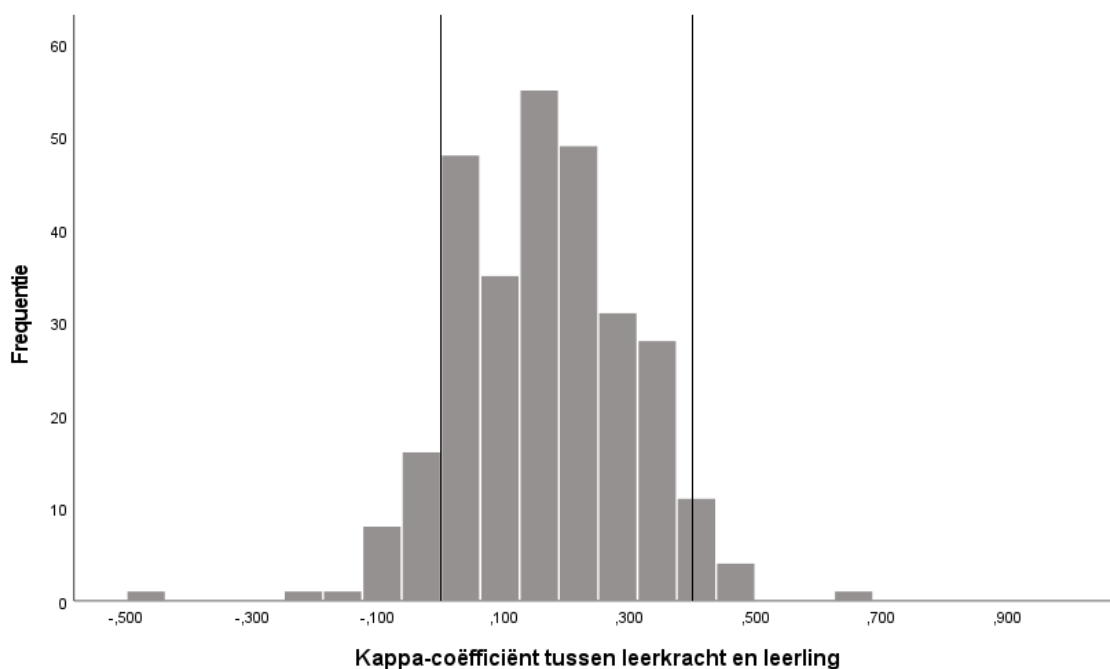
**Figuur 2**

Gewogen Kappa-waarden voor de Overeenstemming tussen Leerling en Ouder, met Afkapscorelijn bij $\kappa < 0$ en $\kappa > .40$



Figuur 3

Gewogen Kappa-waarden voor de Overeenstemming tussen Leerkracht en Leerling, met Afkapscorelijn bij $\kappa < 0$ en $\kappa > .40$



In Tabel 6 is de frequentie gepresenteerd van leerkrachten die betrokken zijn bij minstens vijf keer een matige overeenstemming dan wel onenigheid, waarbij de frequentie van overeenstemming met leerling en ouder is opgeteld. Hierin is te zien dat leerkracht 1, 2 en 3 relatief vaker een matige overeenstemming bereiken, en leerkracht 6, 7 en 8 bereiken relatief vaker onenigheid. De eerste drie leerkrachten zijn samen verantwoordelijk voor 87.76% van het totaal aantal matige overeenstemmingen en de laatste drie leerkrachten zijn samen verantwoordelijk voor 59.46% van het totaal aantal gevallen met onenigheid.

Tabel 6

Frequentie van Leerkrachtbetrokkenheid omtrent Matige Overeenstemming of Onenigheid

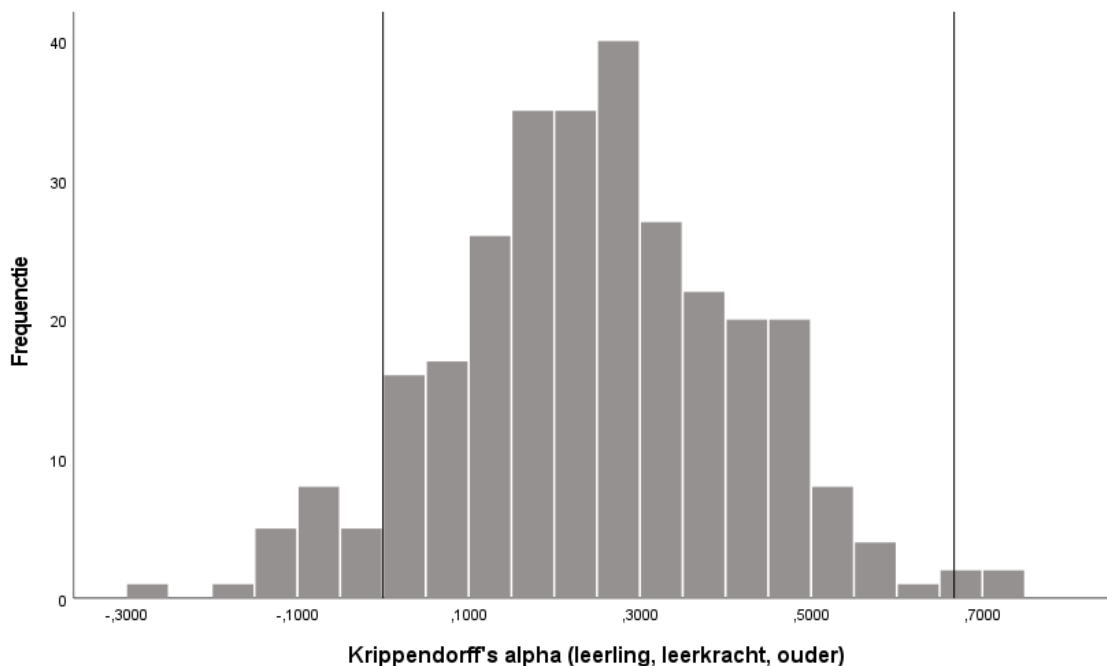
	Matige overeenstemming ($\kappa > .40$)	Onenigheid ($\kappa < 0$)
	<i>n</i> (%)	<i>n</i> (%)
Leerkracht 1	15 (30.61%)	0 (0.00%)
Leerkracht 2	15 (30.61%)	1 (2.70%)
Leerkracht 3	13 (4.13%)	3 (8.11%)
Leerkracht 4	2 (26.53%)	5 (13.51%)
Leerkracht 5	2 (26.53%)	6 (16.22%)
Leerkracht 6	2 (26.53%)	11 (29.73%)
Leerkracht 7	0 (0.00%)	5 (13.51%)
Leerkracht 8	0 (0.00%)	6 (16.22%)

Noot. Percentages zijn berekend op basis van het totaal aantal gevallen binnen de categorie.

De verdeling van de Krippendorff's alfa-waarden, met bijbehorende afkapscorelijn wordt gepresenteerd in Figuur 4. Wanneer gekeken wordt naar de overeenstemming tussen de drie beoordelaars, toont de Krippendorff's alfa dat de overeenstemming over het algemeen laag is ($M = .247$, $SD = .169$, $n = 295$). Voor 20 leerlingen is er systematische onenigheid gevonden ($\alpha_k < 0$), en voor drie leerlingen is de overeenstemming redelijk ($\alpha_k > .667$). Voor de leerlingen met een onenigheid variëren de totaalscores tussen de 245 en 578 en voor de leerlingen met een redelijke overeenstemming liggen de totaalscores tussen de 489 en 540.

Figuur 4

Krippendorff's Alfa tussen Alle Beoordelaars, met afkapscorelijn bij $\alpha_k < 0$ en $\alpha_k > .667$

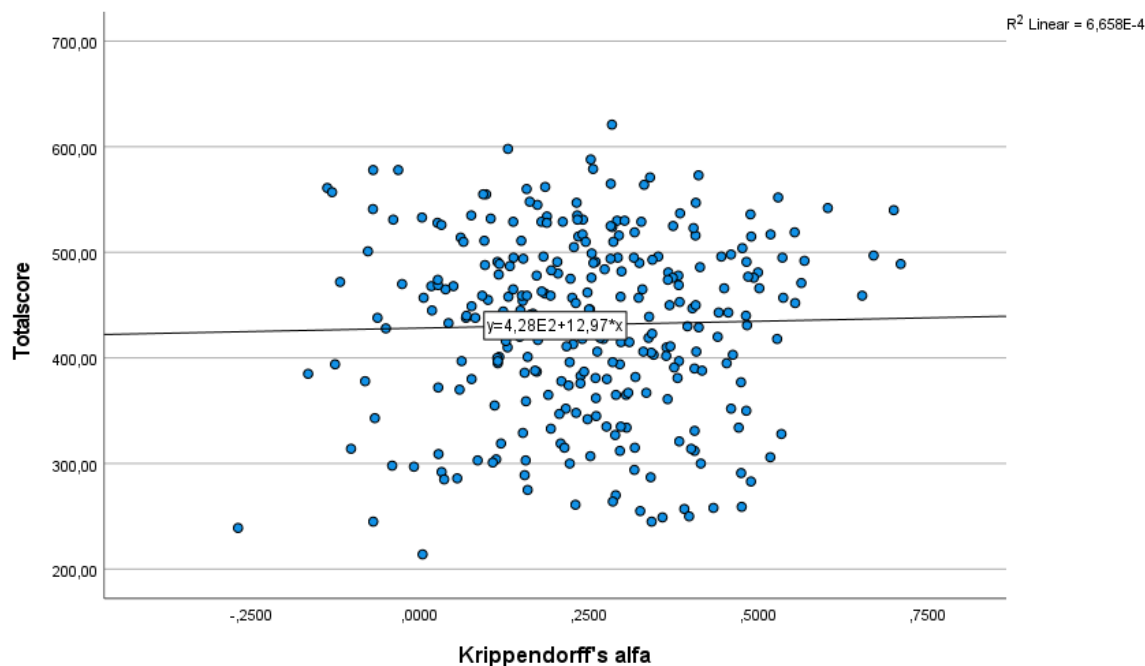


Relatie Interbeoordelaarsbetrouwbaarheid en Scores

Figuur 5 geeft de relatie weer tussen Krippendorff's alfa en de totaalscore van de Vaardighedencheck, waaruit blijkt dat er een zwak positieve, niet-significante, relatie bestaat ($r = .026$, $p = .659$). Dit wil zeggen dat in de steekproef de overeenkomst tussen de drie beoordelaars gemiddeld genomen iets hoger is wanneer de vaardigheden positiever worden beoordeeld.

Figuur 5

Relatie Krippendorff's Alfa en Totaalscore ($n = 295$)



Vergelijking schooljaren

De Shapiro-Wilk test laat zien dat er is voldaan aan de assumptie normaliteit ($p = .731$) en de Levene's test laat zien dat de assumptie van homogeniteit niet geschonden wordt ($p = .942$).

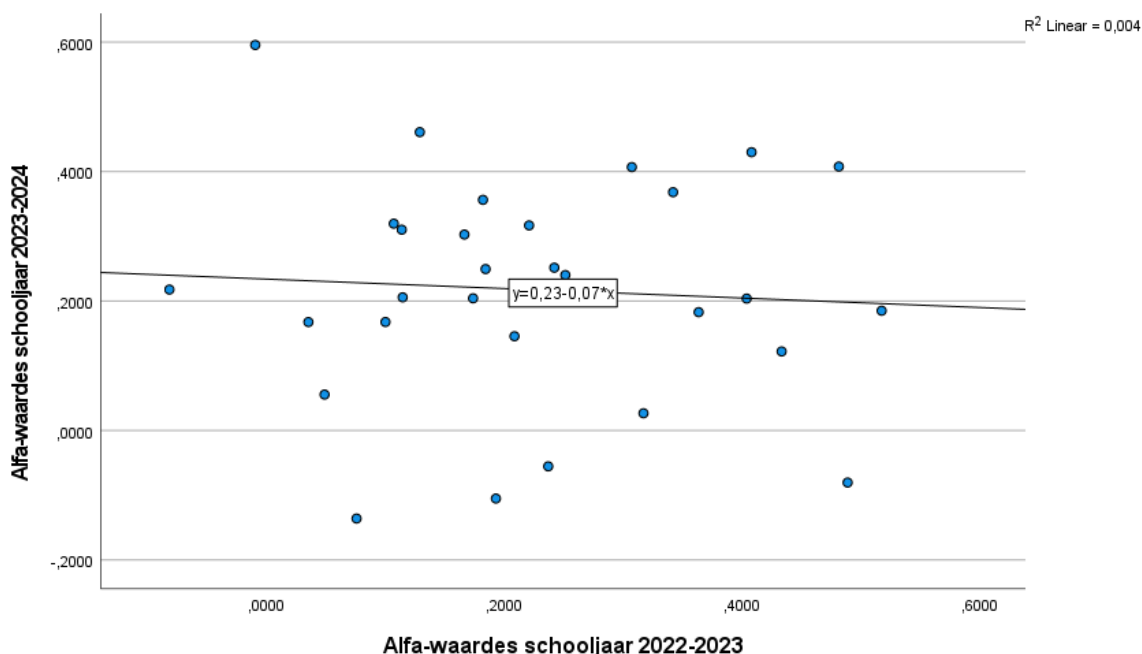
De gemiddelde Krippendorff's alfa-waarde voor het schooljaar 2022-2023 is .226 ($SD = .154$), terwijl de Krippendorff's alfa-waarde voor het schooljaar 2023-2024 .218 ($SD = .174$) bedraagt. Het valt op dat de gemiddelde overeenstemming tussen de drie beoordelaars van het schooljaar 2023-2024 gemiddeld, niet-significant, lager is dan bij de eerste meting, $t(29) = .189$, $p = .851$. In Figuur 6 wordt de relatie van de Krippendorff's alfa van schooljaar 2022-2023 en schooljaar 2023-2024 gepresenteerd, waaruit blijkt dat er een zwak negatief, niet-significant, verband bestaat ($r = -.065$, $p = .733$).

Wanneer er dieper naar de data gekeken wordt, blijkt er in 18 gevallen een gemiddelde hogere overeenstemming bij herhaaldelijk gebruik en in 12 gevallen een lagere overeenstemming bij herhaaldelijk gebruik. Dit wordt verklaard doordat gemiddeld genomen de 18 gevallen met een hogere overeenstemming een kleinere afwijking hebben dan de 12 gevallen met een lagere overeenstemming. Bij deze 12 leerlingen met een lagere algemene overeenstemming na

herhaaldelijk gebruik, blijkt dat in drie gevallen de overeenstemming wel hoger is tussen zowel leerling en leerkracht als tussen leerling en ouder, en in twee gevallen is de overeenstemming wel hoger tussen leerkracht en ouder.

Figuur 6

Relatie Krippendorff's Alfa van Schooljaar 2022-2023 en Schooljaar 2023-2024 ($n = 30$)



Om het effect van een paar extreme observaties te testen, is er een nieuwe Pearson correlatiecoëfficiënt berekend waarbij de twee uiterste observaties zijn weggelaten. Uit deze nieuwe analyse blijkt inderdaad een positieve correlatie tussen de Krippendorff's alfa van schooljaar 2022-2023 en schooljaar 2023-2024 ($r = .193$).

Bij de herhaalde meting waren drie scholen aanwezig met de betrokkenheid van vijf leerkrachten. In Tabel 7 is de frequentie gepresenteerd van de leerkrachtbetrokkenheid omtrent de hogere en lagere algemene overeenstemmingen na herhaaldelijk gebruik. Hieruit blijkt dat twee leerkrachten verantwoordelijk zijn voor 77.78% van de betrokkenheid bij hogere overeenstemmingen en dezelfde leerkrachten zijn verantwoordelijk voor 66.67% van de betrokkenheid bij lagere overeenstemmingen. Verder ontbreekt er voor zes leerlingen structureel data van ouders. In drie van deze gevallen was er een hogere overeenstemming bij herhaaldelijk

gebruik en in de overige drie gevallen was er een lagere overeenstemming na herhaaldelijk gebruik.

Tabel 7

Frequentie van Leerkrachtbetrokkenheid omtrent Hogere en Lagere Krippendorff Alfa-waarde na Herhaaldelijk gebruik

	Hogere overeenstemming	Lagere overeenstemming
Leerkracht 8	7 (38.89%)	3 (25.0%)
Leerkracht 9	7 (38.89%)	5 (41.67%)
Leerkracht 10	2 (11.11%)	2 (16.67%)
Leerkracht 11	1 (5.55%)	2 (16.67%)
Leerkracht 12	1 (5.55%)	0 (0.0%)

Noot. Percentages zijn berekend op basis van het totaal aantal gevallen binnen de categorie.

De items waarbij de algemene overeenstemming het meeste afneemt zijn item 14 (“Voordat ik een opmerking maak, denk ik eerst na”) en item 27 (“Ik kan bedenken welke stappen er nodig zijn om een opdracht te maken”); deze items meten beide een ander construct. Wel blijkt uit de resultaten dat er bepaalde constructen zijn waarbij structurele verschillen zijn. Zo neemt de overeenstemming structureel toe voor de constructen assertiviteit, sociale aansluiting en timemanagement en neemt de overeenstemming structureel af voor de constructen inlevingsvermogen, volgehouden aandacht, taakinitiatie, doorzettingsvermogen en adsorptie. Voor een volledige weergave van alle itemvraagstellingen met bijbehorende constructen, zie Bijlage A, B en C.

Conclusie/Discussie

Conclusie

Het doel van het huidige onderzoek is om de overeenstemming in opvattingen tussen leerling, ouder en leerkracht te analyseren, met als bredere doelstelling het verfijnen van de Vaardighedencheck. De eerste onderzoeksvraag die in dit onderzoek centraal staat is: “In hoeverre komen de opvattingen over de vaardigheden van leerling, ouder en leerkracht, zoals gemeten in de Vaardighedencheck, met elkaar overeen?”. Over het algemeen blijkt dat de overeenstemming tussen de drie beoordelaars beperkt is. De overeenstemming tussen ouder en leerkracht is gemiddeld genomen redelijk en lijkt daarmee iets sterker dan zowel de overeenstemming tussen leerling en ouder, als tussen leerling en leerkracht, waarbij de overeenstemming gemiddeld genomen zwak is. De gevonden afwijkingen tussen leerling en leerkracht lijken niet structureel te zijn, terwijl leerlingen over het algemeen wel een positievere beoordeling toekennen aan zichzelf. De afwijkingen tussen zowel leerling als leerkracht met ouders lijkt wel structureel, waarbij ouders vaak een negatievere beoordeling geven.

Bij nader inzien van de data blijkt dat er enkele uitzonderingen zijn op de algemene beperkte overeenstemming tussen de drie beoordelaars. Opvallend is dat de drie leerlingen met een redelijke algemene overeenstemming, ook relatief hogere totaalscores behalen op de Vaardighedencheck. Ook zijn er enkele uitzonderingen op de algemene zwakke tot redelijke overeenstemming tussen de verschillende beoordelaarsparen. In de gevallen met een matige overeenstemming valt op dat drie leerkrachten elk verantwoordelijk zijn voor meer dan een kwart van deze beoordelingen, wat samen bijna 88% van alle matige overeenstemmingen vertegenwoordigt. In de gevallen waarin sprake is van onenigheid, blijkt één leerkracht meer dan een kwart van de verantwoordelijkheid te dragen. Dit duidt erop dat sommige leerkrachten vaker tot een hogere mate van overeenstemming komen, terwijl anderen juist vaker betrokken zijn bij uiteenlopende beoordelingen. Aangezien de grote invloed van deze enkele leerkrachten, valt het niet uit te sluiten dat de conclusies van het onderzoek anders zouden zijn wanneer deze desbetreffende leerkrachten niet in de steekproef zouden zijn opgenomen.

Over het algemeen zijn er weinig items waarbij de opvattingen tussen beoordelaars structureel afwijken. Voor zowel de beoordeling tussen leerkracht en ouder en tussen leerkracht

en leerling, wijkt er slechts één item structureel af. Terwijl er tussen de beoordeling van leerling en ouder zes items structureel afwijken, waarbij in vijf van deze gevallen de leerling positiever is.

De tweede hoofdvraag van het onderzoek luidt: “In hoeverre komen de, in de Vaardighedencheck gemeten, opvattingen over de vaardigheden van leerling/ouder/leerkracht dichter bij elkaar na een bespreking van de Vaardighedencheck?” De opvattingen blijken gemiddeld genomen niet dichter bij elkaar te komen na een bespreking. De gemiddelde overeenstemming tussen de drie beoordelaars voor het schooljaar 2022-2023 is zelfs hoger dan de gemiddelde overeenstemming tussen de drie beoordelaars voor het schooljaar 2023-2024. Echter, dit is een klein en niet-significant verschil, gezien het gaat om een verschil van $\alpha .008$, waarbij een waarde van $\alpha .667$ pas wordt gezien als redelijke overeenstemming. Opvallend is dat er in het merendeel sprake was van een iets hogere algehele overeenstemming, maar doordat de gevallen met een lagere algehele overeenstemming sterker afwijken, blijkt de gemiddelde algehele overeenstemming na bespreking uiteindelijk toch lager te liggen. Het is niet altijd het geval dat alle drie de beoordelaars het minder eens zijn met elkaar na een bespreking. In deze gevallen bleek voor soms de overeenstemming tussen twee van de drie beoordelaars wel hoger te zijn. Dit kwam niet opvallend vaker voor bij een specifiek beoordelaarspaar.

Wanneer we kijken naar de leerkrachtbetrokkenheid van de herhaalde metingen, valt het op dat twee leerkrachten een groot aandeel leveren in de resultaten. Deze twee leerkrachten zijn verantwoordelijk voor meer dan driekwart van de betrokkenheid bij een hogere algehele overeenstemming en tegelijk zijn zij ook verantwoordelijk voor een derde van de betrokkenheid bij een lagere algehele overeenstemming. Dit duidt er niet op dat sommige leerkrachten vaker tot een hogere overeenstemming komen dan anderen, maar wel dat sommige leerkrachten een grote invloed hebben op de steekproef. Het valt daarom wederom niet uit te sluiten dat de conclusies van het onderzoek anders zouden zijn wanneer deze desbetreffende leerkrachten niet in de steekproef zouden zijn opgenomen.

Voor de meeste items en constructen werd er geen structurele afwijking tussen de algehele overeenstemming gevonden. Van de vijf constructen waarbij het een structurele afname betreft, gaan drie constructen over executieve vaardigheden.

Dat ouders vaak een negatievere beoordeling geven komt overeen met de eerder benoemde literatuur. Uit onderzoek van Murray et al. (2011) bleek namelijk dat ouders meer risico's ervaren dan leerkrachten. Ook de bevinding dat leerlingen zichzelf positiever beoordelen is, sluit aan bij de literatuur. Uit het Dunning-Kruger-effect blijkt dat (minder competente) mensen vaak hun eigen vaardigheden overschatten (Dogan et al., 2023; Hoger et al., 2022), specifiek met een dip aan het begin van de pubertijd (Crone & Drunen, 2024). Dit kan tevens een verklaring zijn voor de gevonden hogere overeenstemming tussen het beoordelaarspaar leerkracht-ouder, waarvan de leerling zelf geen onderdeel uitmaakt. Een mogelijke verklaring voor de redelijke overeenstemming bij leerlingen met een hogere totaalscore ligt mogelijk in het 'halo-effect'. Dit effect beschrijft dat mensen met een aantal positieve eigenschappen anderen ertoe kunnen aanzetten hen te beoordelen met meer positieve eigenschappen, wat kan leiden tot homogene perceptie (Jinzhi, 2024; Sanrey et al., 2021). Volgens deze theorie kan het zijn dat een leerling met een aantal sterke vaardigheden als geheel vaardig wordt ervaren, wat leidt tot een grotere algehele overeenstemming bij een hogere totaalscore. Hoewel dit mogelijk een verklaring kan zijn voor het verschil in overeenstemming, moet dit voorzichtig geïnterpreteerd worden aangezien de bevinding slechts geldt voor drie leerlingen. Wel sluit het ook aan bij de gevonden zwakke positieve relatie tussen de Krippendorff's alfa en de totaalscore. Verder blijkt uit onderzoek van Spreij et al. (2024) dat de opvattingen over executieve vaardigheden duidelijk verschillen tussen leerling, ouder en leerkracht. Van de items met structurele gevonden verschillen, gaat het merendeel over executieve en schoolse vaardigheden van de leerling. Ook bij de herhaalde metingen is gevonden dat het merendeel van de structurele afnames executieve functies betreft. De beoordeling van executieve functies is gevoelig voor subjectieve waarnemingen. Een mogelijke verklaring voor de opvattingsverschillen over deze vaardigheden tussen leerkracht en ouder is dat ze het kind in een andere omgeving zien en daardoor een andere interactie hebben met de desbetreffende leerling (Bergström & Baviskar, 2021; Toplak et al., 2013). Daarnaast is de vraag in hoeverre een ouder in de Vaardighedencheck een oordeel kan geven over executieve vaardigheden, zoals bijvoorbeeld item 1 ("Mijn kind stelt tijdens de les vragen aan klasgenoten"). De verschillende opvattingen bieden in ieder geval een aanknopingspunt tijdens een driehoeksgesprek.

Het feit dat de algehele overeenstemming gemiddeld genomen lager is na herhaaldelijk gebruik gaat in tegen de verwachting van de onderzoekers. Zoals eerder aangegeven signaleren ouders mogelijk meer risico's (Murray et al., 2011) en overschatten leerlingen zichzelf vaak

(Hoger et al., 2022). De aanname is dat bij herhaaldelijk gebruik van de Vaardighedencheck er tussendoor een bespreking is geweest, in de vorm van een driehoeksgesprek. Het is mogelijk dat een driehoeksgesprek deze kloof niet gelijk dicht. Daarnaast is het denkbaar dat er tijdens het driehoeksgesprek misverstanden ontstaan of dat er sprake is van gebrekkige communicatie. Het hebben van een gemeenschappelijk perspectief en het hebben van een gedeelde taal is belangrijk bij de samenwerking tussen school en ouders (Wienen et al., 2020).

Discussie

Het huidige onderzoek toont enkele methodologische tekortkomingen die invloed hebben op de validiteit van de resultaten. Ten eerste wordt de interne validiteit besproken. Er is bekend geworden dat de woordkeuze in de Vaardighedencheck bij zeven items niet volledig gelijk is voor de drie beoordelaars (zie Bijlage A, B en C). Mogelijk leidt dit verschil in formulering tot een verzwakking van de constructvaliditeit. Door de verschillende woordkeuzes kunnen beoordelaars verschillende interpretaties hebben van de items. Daarnaast wordt aangenomen dat het voeren van een driehoeksgesprek heeft plaatsgevonden, maar het is niet duidelijk of dit in de praktijk daadwerkelijk is gebeurd. Het ontbreken van inzicht in de uitvoering van dit gesprek kan de interne validiteit ondermijnen, omdat het effect van deze aannames mogelijk niet correct wordt geëvalueerd. Daarnaast werd een kwart van de items niet ingevuld door ouders, voornamelijk doordat volledige vragenlijsten niet ingevuld werden. Deze missende data zijn mogelijk van invloed op de interne validiteit, doordat de overeenstemming met het beoordelingspaar ouder vertekend is. Door ontbrekende informatie kan niet gesteld worden of deze missende data willekeurig is, of dat het bijvoorbeeld gaat om een systematisch patroon, zoals bijvoorbeeld bepaalde type ouders die de Vaardighedencheck niet invulden. Desondanks wijst de uitgevoerde *t*-toets uit dat er geen structureel verschil bestaat tussen de groepen met en zonder missende data, wat suggereert dat de missende data geen significante invloed hebben op de interne validiteit binnen dit onderzoek. Wegens de grootte van de steekproef wordt aangenomen dat de missende data weinig effect hebben op de conclusies van het onderzoek.

Ten tweede wordt de externe validiteit besproken. Het hanteren van een gemakssteekproef kan een negatieve invloed hebben op de externe validiteit. Het gebruik van een gemakssteekproef kan namelijk leiden tot een steekproef die minder representatief is voor de beoogde doelpopulatie (Merkus, 2023). In dit onderzoek kan een gemakssteekproef er bijvoorbeeld voor hebben gezorgd

dat scholen alleen de Vaardighedencheck hebben ingezet bij zorgleerlingen en/of wanneer er een lage overeenstemming verwacht wordt tussen leerkracht/ouder/leerling. Hierover is geen aanvullende informatie beschikbaar, wat het vaststellen van mogelijke effecten op de conclusies van dit onderzoek lastig maakt. Tegelijkertijd werd echter beoogd de externe validiteit te vergroten door de data te verzamelen in een natuurlijke setting en de variabelen enkel te observeren zonder dat er sprake was van manipulatie en/of toewijzing aan condities. Ook de grote steekproefomvang heeft een positieve bijdrage aan de externe validiteit. Bij het beantwoorden van de eerste onderzoeksvraag werd gebruikgemaakt van een grote steekproef; 240-295 deelnemers. Over het algemeen wordt een minimale steekproefgrootte van 149 als voldoende beschouwd bij het uitvoeren van correlatieanalyses (Bujang, 2024). Hierdoor wordt verwacht dat de conclusies van dit onderzoek op zijn minst generaliseerbaar zijn in de regio Twente. Generaliseerbaarheid naar de rest van Nederland is minder stellig, maar wel mogelijk gezien de overeenkomst in schoolsysteem; wat in Nederland landelijk geregeld is (Ministerie van Algemene Zaken, 2024). Daarentegen is voor het beantwoorden van de tweede deelvraag slechts data van een beperkt aantal deelnemers ($n = 30$) geanalyseerd. Ondanks de beperkte omvang van deze tweede steekproef is voldaan aan de assumpties voor het uitvoeren van een statistische toets, wat het mogelijk heeft gemaakt om resultaten van een statistische toets te verkrijgen. Niettemin zorgt de beperkte steekproefomvang ervoor dat enkele observaties al een groot effect hebben, wat leidt tot minder zekerheid van de conclusies. Hierdoor is de generaliseerbaarheid van de conclusie op de tweede hoofdvraag minder sterk. Om het effect van een paar extreme observaties te testen, is er een nieuwe Pearson correlatiecoëfficiënt berekend, waaruit blijkt dat zonder de twee uiterste observaties de correlatie tussen de Krippendorff's Alfa van Schooljaar 2022-2023 en Schooljaar 2023-2024 positief is. Dit toont aan dat de conclusies van het onderzoek anders waren geweest wanneer deze deelnemers niet in de steekproef waren opgenomen. Het is dus belangrijk om bewust te zijn van de grote impact die de uitschieters hebben op de resultaten van het onderzoek. Ondanks dat de conclusie voor de tweede hoofdvraag minder zeker is, lijkt het niet vertekend. In de herhaalde meting waren drie scholen aanwezig met de betrokkenheid van vijf leerkrachten. Hierbij is geen structureel verschil gevonden wat betreft de mate van overeenstemming tussen de drie beoordelaars; de steekproef lijkt dus aselekt, wat positief is voor de externe validiteit.

Gezien de beperkte overeenstemming tussen de drie beoordelaars, wordt op basis van het huidige onderzoek aanbevolen om de drieperspectievenmeting voort te zetten bij het gebruik van de Vaardighedencheck. Eerder onderzoek stelt ook dat het belangrijk is om vanuit verschillende perspectieven dezelfde vaardigheden te meten; vooral de beoordeling door ouders wordt als belangrijk ervaren (Bergstorm & Baviska, 2021; Vugteveen, 2021). Binnen huidig onderzoek blijkt dat de beoordeling van leerling zelf het minst overeenstemming oplevert, waardoor het belangrijk is om dit perspectief te blijven gebruiken. Bovendien blijkt actieve betrokkenheid en zelfbeoordeling een positief effect te hebben op het verbeteren van het eigen leerproces van leerlingen (Boud & Falchikov, 2006) en draagt het bij aan het positief welzijn van leerlingen en heeft het een positieve invloed op toekomstige leerprestaties (Jindal-Snape et al., 2019; Korpershoek et al., 2016; Van Rens, 2020). Het vroegtijdig belichten van verschillende opvattingen door middel van de Vaardighedencheck kan het adviesgesprek in groep 8 vergemakkelijken en zorgen voor een gedeelde mening over het uitstroomperspectief van de leerling. Bewustwording van patronen kan hierbij al helpend zijn. Een training voor leerkrachten om meer inzicht te krijgen over structurele verschillen in perceptie van zichzelf, de leerling en de ouder kan mogelijk voor erkenning en relativering zorgen. Daarnaast hoeven verschillen in opvatting niet per definitie negatief te zijn, maar kan dit ook kansen bieden om de vaardigheden van de leerling vanuit verschillende invalshoeken te belichten. Momenteel wordt door de ontwikkelaars van de Vaardighedencheck de keuze gemaakt om een totaaloverzicht weer te geven waarin vaardigheden geordend worden op basis van de gemiddelde score van de drie beoordelaars (Van Gennip & Klinkenberg, 2023). Op deze manier wordt er wellicht weinig aandacht besteed aan de variantie in opvatting tussen de beoordelaars, terwijl daar juist het belang van een bespreking ligt. Ordening zou ook gedaan kunnen worden op basis van de items met de meeste verschillen in opvatting.

Over het algemeen is de overeenstemming tussen leerling en ouder het laagst. Een school kan meerdere mogelijkheden kunnen inzetten om dit verschil in perceptie te verkleinen. Scholen kunnen bijvoorbeeld meeloopdagen organiseren, waarbij ouders de schoolvaardigheden en het gedrag van hun kinderen in de onderwijspraktijk kunnen observeren. Daarnaast kunnen mentoren een bemiddelende rol vervullen tussen leerling en ouder, waarbij zij op een neutrale manier beide perspectieven bespreekbaar maken.

Verschil in opvatting tussen de drie beoordelaars komt niet voor elk item in dezelfde mate voor en er zijn enkele items met structurele verschillen tussen de beoordelaars. Bij nader inzien van de vraagstellingen voor desbetreffende items valt op dat bij meerdere items de verwoording voor de leerling (zie Bijlage A) niet hetzelfde is als de verwoording voor de leerkracht/ouder (zie Bijlage B en C). Een advies is daarom om de verwoording bij alle items consistent te maken voor alle drie de beoordelaars. Voor de overige items, waarbij structurele verschillen zijn gevonden maar de verwoording wel volledig overeenkomt, wordt aangeraden om extra toelichting te geven. Dit om uit te kunnen sluiten dat de gevraagde vaardigheid duidelijk is. Uit literatuur blijkt dat het gebruik van goed geformuleerde items een positief effect heeft op de interbeoordelaarsbetrouwbaarheid. Duidelijk geformuleerde items verminderen namelijk ruis (Van der Ark & ten Hove, 2019). Ook is het mogelijk dat de items wel duidelijk zijn voor de beoordelaars, maar dat er toch verschil in opvatting blijft bestaan. In dat geval is het belangrijk voor leerkrachten om deze items meer aandacht te geven tijdens een driehoeksgesprek.

Dat de overeenkomst tussen de drie beoordelaars niet dichterbij elkaar komt na een bespreking gaat in tegen de verwachting van de onderzoekers. Tijdens een driehoeksgesprek is het van belang dat er een gezamenlijk doel komt en dat leerling, ouder en leerkracht tot meer overeenstemming komen. Wanneer het voeren van een driehoeksgesprek niet de gewenste effecten met zich meebrengt, kan het inzetten van een neutrale vierde persoon, zoals een intern begeleider of gedragswetenschapper, wellicht een oplossing bieden.

Het is van belang dat er vervolgonderzoek wordt uitgevoerd naar de Vaardighedencheck, aangezien er op verschillende onderdelen nog kennis ontbreekt. Voor vervolgonderzoek is het belangrijk om een grotere steekproef te werven van beoordelaars die de Vaardighedencheck zowel voor als na een driehoeksgesprek invullen. Dit zou bijdragen aan inzicht in de mate waarin opvattingen van leerling, ouder en leerkracht naar elkaar toe groeien na een bespreking. Daarnaast zou het ook interessant zijn om meer informatie te verzamelen, zoals ervaring van leerkracht of opleidingsniveau van ouder, om te kunnen onderzoeken welke factoren van invloed zijn op de mate van overeenstemming tussen de beoordelaars en waarom perspectieven soms juist verder uit elkaar lopen na een bespreking. Ook extra informatie over de onderlinge relatie tussen beoordelaars zou relevant zijn. De indruk die men van iemand heeft, heeft namelijk invloed op hoe men diegene beoordeelt (Humphrey-Murto et al., 2021). Een leerling die een goede band heeft met

zijn leerkracht en/of ouder komt mogelijk tot een hogere overeenstemming dan een leerling die dat niet heeft.

Daarnaast is het wenselijk om na te gaan in hoeverre beoordelaars dezelfde interpretatie hebben over de gevraagde vaardigheden. Dit kan bijvoorbeeld worden onderzocht door beoordelaars hardop te laten denken tijdens het invullen van de Vaardighedencheck of door aansluitend aan het invullen van de Vaardighedencheck een interview af te nemen.

Tot slot biedt een longitudinale onderzoeksopzet mogelijkheden om leerlingen te volgen, om zo te onderzoeken of de Vaardighedencheck voorspellende waarde heeft voor uitkomsten zoals schoolsucces of schoolwelzijn in het voortgezet onderwijs.

De resultaten van dit onderzoek laten zien dat de overeenstemming tussen leerling, ouder en leerkracht over het algemeen beperkt is. Juist daarom is het van belang voor scholen om de verschillen in opvatting vroeg te signaleren. Dit bevordert open communicatie en kan het adviesgesprek vergemakkelijken, met het doel om tot een gezamenlijk uitstroomperspectief te komen. Opvallend is dat de driehoeksgesprekken gemiddeld genomen niet leiden tot een hogere mate van overeenstemming tussen leerling, ouder en leerkracht. De bevindingen van dit onderzoek onderstrepen de meerwaarde van een drieperspectievenmeting en roepen tegelijkertijd vragen op over de effectiviteit van bestaande gespreksvormen. Naar voren komt de noodzaak voor verdiepend onderzoek; zowel naar de ontwikkeling van de Vaardighedencheck als naar effectieve interventies waarmee scholen de perspectieven dichter bij elkaar kunnen brengen. Het uiteindelijke doel blijft duidelijk: in een goede flow naar het VO.

Literatuurlijst

- Bergström, M., & Baviskar, S. (2020). A systematic review of some reliability and validity issues regarding the Strengths and Difficulties Questionnaire focusing on its use in out-of-home care. *Journal of Evidence-Based Social Work*, 18(1), 1–31.
<https://doi.org/10.1080/26408066.2020.1788477>
- Besi, M., & Sakellariou, M. (2019). Factors associated with the successful transition to primary school. *European Journal Of Education Studies*. <https://doi.org/10.5281/zenodo.2555431>
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31(4), 413–426.
- Bujang, M. A. (2024). An elaboration on sample size determination for correlations based on effect sizes and confidence interval width: a guide for researchers. *Restorative Dentistry & Endodontics*, 49(2). <https://doi.org/10.5395/rde.2024.49.e21>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.
<https://doi.org/10.1037/h0026256>
- Crone, E. A., & van Drunen, L. (2024). Development of self-concept in childhood and adolescence: How neuroscience can inform theory and vice versa. *Human development*, 68(5-6), 255–271. <https://doi.org/10.1159/000539844>
- Dogan, E., Güven, S., Dal, N., Tok, S., & Isik, U. (2023). An Investigation of the Dunning-Kruger Effect in First-Year College Students during Distance Education. *Journal of Educators Online*, 20(4).
- Driessen, G., Doesborgh, J., Ledoux, G., Overmaat, M., Roeleveld, J., & Van der Veel, I. (2005). De overgang naar het voortgezet onderwijs. Nijmegen/Amsterdam: ITS/SCO Kohnstamm Instituut.
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, 82(1), 405–432.
<https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Evans, D., Borriello, G. & Field, A. (2018) A review of the academic and psychological impact of the transition to secondary education. *Frontiers in Psychology*, 9, art. 1482, 1-18.
- Harris, J. & Nowland, R. (2021) Primary-secondary school. Transition: Impacts and

- opportunities for adjustment. *Journal of Education and Social Sciences*, 8 (2). pp. 55-69. ISSN 2414-8091.
- Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods And Measures*, 1(1), 77–89.
<https://doi.org/10.1080/19312450709336664>
- Hofer, G., Mraulak, V., Grinschgl, S., & Neubauer, A. C. (2022). Less-intelligent and unaware? Accuracy and Dunning-Kruger effects for self-estimates of different aspects of intelligence. *Journal of Intelligence*, 10.
- Humphrey-Murto, S., Shaw, T., Touchie, C., Pugh, D., Cowley, L., & Wood, T. J. (2021). Are raters influenced by prior information about a learner? A review of assimilation and contrast effects in assessment. *Advances in Health Sciences Education*, 26(3), 1133–1156.
<https://doi.org/10.1007/s10459-021-10032-3>
- Jindal-Snape, D., Symonds, J., Hannah, E. & Barlow, W. (2021). Conceptualising primary-secondary school transitions: A systematic mapping review of worldviews, *Theories and Frameworks. Frontiers in Education*, 6, art. 540027, 1-15.
- Jinzhi, H. (2024). Halo effect. *The ECPH Encyclopedia of Psychology*.
<https://doi.org/10.1007/978-981-97-7874-4>
- Koren, L. (2024, 7 maart). *Hoe komt het schooladvies tot stand in groep 8?* Onderwijsconsument. Geraadpleegd op 30 november 2024, van
<https://www.onderwijsconsument.nl/hoe-komt-het-basisschooladvies-tot-stand>
- Korpershoek, H., Beijer, C., Spithoff, M., Naaijer, H. M., Timmermans, A.C., Van Rooijen, M., Vugteveen, J., & Opdenakker, M.-C. (2016). Overgangen en aansluitingen in het onderwijs. Deelrapportage 1: reviewstudie naar de po-vo en de vmbo-mbo overgang. *GION onderzoek/onderwijs*.
- Krippendorff, K. (2004). Content analysis: An introduction to its methodology (2de editie). *SAGE publications*.
- Marasini, D., Quatto, P., & Ripamonti, E. (2014). Assessing the inter-rater agreement for ordinal data through weighted indexes. *Statistical Methods in Medical Research*, 25(6), 2611–2633. <https://doi.org/10.1177/0962280214529560>
- Merkus, J. (2023). Een selecte of aselechte steekproef in je scriptie. *Scribbr*. Geraadpleegd op 4 april 2025, van <https://www.scribbr.nl/onderzoeksmethoden/steekproef-in-je-scriptie/>

- Naaijer, H. M., Spithoff, M., Osinga, M., Klitzing, N., Korpershoek, H., & Opdenakker, M.-C. (2016). De overgang van primair naar voortgezet onderwijs in internationaal perspectief: Een systematische overzichtsstudie van onderwijstransities in relatie tot kenmerken van verschillende Europese onderwijsstelsels. *GION onderwijs/onderzoek*.
<https://www.nro.nl/kb/405-15-724-overgang-van-primair-naar-voortgezet-onderwijs-ininternationaal-perspectief/>
- Rapportage Schooladviezen. (2020). In *duo-onderwijsonderzoek.nl*. DUO onderwijsonderzoek & advies. Geraadpleegd op 11 november 2024, van <https://www.duo-onderwijsonderzoek.nl/wp-content/uploads/2020/02/Onderzoekschooladviezen-11-februari-2020.pdf>
- Sanrey, C., Bressoux, P., Lima, L., & Pansu, P. (2021). A new method for studying the Halo Effect in teachers' judgement and its antecedents: Bringing out the role of certainty. *British Journal of Educational Psychology*, 91(2), 658–675. <https://doi.org/10.1111/bjep.12385>
- Smith, J.S., Akos, P., Lim, S., & Wiley, S. (2008). Student and stakeholder perceptions of the transition to high school. *The High School Journal* 91(3), 32-42.
<https://dx.doi.org/10.1353/hsj.2008.0003>
- Spreij, L. C., Van Tuijl, C., & Leseman, P. P. M. (2024). Validating rating scales for executive functioning across education levels and informants. *Contemporary School Psychology*.
<https://doi.org/10.1007/s40688-023-00462-8>
- Van der Ark, L. A., & ten Hove, D. (2019). Zijn we het eens? Interbeoordelaarsbetrouwbaarheid in de pedagogiek en het onderwijs. *Pedagogische Studiën*, 95(5/6), 361-371.
<http://pedagogischestudien.nl/search?identifieer=681917>
- Van Gennip, R., & Klinkenberg, M. (2023). Rapportage evaluatieonderzoek: Herinrichting van de inhoud van de commissies 10-14.
- Van Rens, M., Haelermans, C., Groot, W., & Maassen van den Brink, H. (2017). Facilitating a successful transition to secondary school: (how) does it work? A systematic literature review. *Adolescent Research Review*, 3(1), 43-56.
- Van Rens, M. (2020). *The transition from primary to secondary education: a study of the voices of children and teachers in the Netherlands*. <https://doi.org/10.26481/dis.20200904mr>

- van Rooijen, M., Korpershoek, H., Vugteveen, J., & Opdenakker, M. C. (2017). De overgang van het basis- naar het voortgezet onderwijs en de verdere schoolloopbaan. *Pedagogische Studiën*, 94(2), 110- 134. <http://pedagogischestudien.nl/search?identifier=632153>
- Warrens, M. J. (2017). Symmetric kappa as a function of unweighted kappas. *Communications in Statistics. Part B: Simulation and Computation*, 46(7), 5240-5245. <https://doi.org/10.1080/03610918.2016.1146765>
- Wienen, A., Korenromp, H., Batstra, L., & Graas, T. A. M. (2020). NRO rapport. Werkbaar in de klas: Het effect van samenwerking tussen ouders, leraren, maatschappelijk werk en jeugdhulp in de schoolklas. *Nationaal Regieorgaan Onderwijsonderzoek*.
- Yu, T., Yang, L., Jiang, X., Shao, S., Sha, W., & Li, M. (2023). Matters arising: methodological issues on evaluating agreement between medical students' attitudes towards drugs/alcohol use during pregnancy by Cohen's kappa analysis. *BMC Medical Education*, 23(1). <https://doi.org/10.1186/s12909-023-04071-0>