



Samenvatting Artikelen A&E 2021

VOCUS heeft deze samenvatting te danken aan Melanie Kamp.

Het gebruik van deze samenvatting is bedoeld als studeerhulp na het lezen van de verplichte literatuur. Gebruik van deze samenvatting is geheel voor eigen risico.

Soms wordt er verwezen naar bladzijden of tabellen in het originele boek.

Succes met studeren!



Assessment & Evaluatie

Inhoud

1.1	Poorthuis et al. (2014) Do grades shape student' school engagement? The psychological consequences of report card grades at the beginning of secondary school	3
1.2	Black & William (2018) Classroom assessment and pedagogy.....	5
1.3	Gerritsen et al. (2017) Assessment quality in tertiary education: An integrative literature review	8
2.1	Howard et al. (2018) What are standardised literacy and numeracy test testing? Evidence of the domain general contributions to students' standardised educational test performance.....	11
2.2	Butler (2018) Multiple choice testing in education: Are the best practices for assessment also good for learning?	13
2.3	Brookhart (2018) Appropriate criteria: Key to effective rubrics	14
6.1	Van der Gaag et al. (2019) Evidence for measuring teachers' core practices	16
6.2	O'Connel et al. (2016) Does calibration reduce variability in the assessment of accounting learning outcomes?	17
6.3	Bok et al. (2018) Validity evidence for programmatic assessment in competency-based education.....	19
7.1	Boud & Soler (2016) Sustainable assessment revisited.....	20
7.2	Panadero et al. (2017) Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analysis	22
7.3	Carless & Boud (2018) The development of student feedback literacy: enabling uptake feedback.....	25

1.1 Poorthuis et al. (2014) Do grades shape student' school engagement? The psychological consequences of report card grades at the beginning of secondary school

Abstract:

Lagere rapportcijfers voorspellen een lagere emotionele en gedragsmatige betrokkenheid. Jongens en kinderen die de prestatienormen in hun rapporten als hoog beschouwden, reageerden meer affectief op hun cijfers, wat resulteerde in een sterker indirect effect van cijfers via negatief affect op emotionele betrokkenheid.

Introduction:

Cijfers en de affectieve reacties die zij oproepen, beïnvloeden waarschijnlijk hoeveel tijd en moeite studenten in hun latere werk investeren. Deze studie stelt voor dat de affectieve reacties van studenten op cijfers zouden kunnen helpen verklaren hoe en waarom cijfers een invloed hebben op de latere betrokkenheid van studenten.

School Engagement

Schoolbetrokkenheid (school engagement) - de actieve betrokkenheid van leerlingen bij schoolse activiteiten - is een van de sterkste voorspellers van slagen of falen op school.

- Behavioral engagement: verwijst naar openlijk gedrag, zoals actieve klasdeelname en aandacht
- Emotional engagement: heeft betrekking op psychologische ervaringen, zoals belangstelling en enthousiasme

Role of Affective reactions

Motivatietheorieën benadrukken de rol van cognitieve processen als voorspellers van schoolbetrokkenheid. Cijfers voorspellen schoolbetrokkenheid door de positieve en negatieve affectieve reacties

- Positief affect wordt gekenmerkt door het opwekken van aangename en alerte gemoedstoestanden, zoals opwinding. Gebrek aan positief affect wordt gekenmerkt door lusteloosheid
- Negatieve affecten wekken aversieve stemmingen op, zoals verdriet en schaamte. Gebrek aan negatief affect wordt gekarakteriseerd door kalmte. Negatief affect kan ertoe leiden dat studenten zich psychologisch terugtrekken, bijvoorbeeld doordat zij zich minder inspannen voor hun schoolwerk. Negatief affect is in verband gebracht met een lagere behavioral betrokkenheid. Negatief affect kan sommige leerlingen echter motiveren om een aversieve toestand te overwinnen en toekomstige mislukking te voorkomen door extra inspanning in hun schoolwerk te investeren. Soms moedigt het mensen aan om harder hun best te doen.

Moderating factors: Performance norms and gender

Leerlingen ontvangen hun cijfers niet in een sociaal vacuüm, maar zullen hun prestaties waarschijnlijk vergelijken met die van hun klasgenoten. De perceptie die leerlingen hebben van de prestatienormen in hun klas voorspelt hun latere gedrag. Rapportkaarten kunnen sterkere affectieve reacties oproepen bij leerlingen die de normen in hun klas als hoog inschatten dan bij leerlingen die de normen als laag inschatten.

In vergelijking met jongens halen meisjes doorgaans hogere schoolcijfers en zijn ze ook meer betrokken bij de school. Het kan zijn dat jongens zwakkere affectieve gevoelens ervaren dan meisjes als reactie op hun schoolprestaties. Maar jongens zijn competitiever en het kan dus ook zijn dat zij sterkere affectieve reacties op cijfers ervaren.

Present study

1. Het eerste doel van de studie was na te gaan of de resultaten van de schoolkaarten van de leerlingen de latere schoolbetrokkenheid voorspellen.
2. Het tweede doel was na te gaan of het veronderstelde verband tussen rapportcijfers en schoolbetrokkenheid wordt bevorderd door de positieve en negatieve affectieve reacties die cijfers oproepen.
3. Het derde doel was na te gaan of de veronderstelde bemiddeling van positief en negatief affect wordt gemedereerd door de perceptie die leerlingen hebben van de prestatienormen in hun klas.

Results

Zoals verwacht, waren rapportcijfers gerelateerd aan zowel positieve als negatieve affectieve reacties op cijfers en aan emotionele en gedragsmatige betrokkenheid.

Er waren twee sekseverschillen

- meisjes behaalden hogere cijfers dan jongens
- meisjes ervoeren minder negatief affect in reactie op hun cijfers dan jongens
- geen verschil in positief affect

Emotionele betrokkenheid nam af in de loop van het eerste schooljaar, waaruit blijkt dat emotionele betrokkenheid sterker afnam bij jongens dan bij meisjes. Er werden geen sekseverschillen gevonden voor gedragsmatige betrokkenheid.

Rapportkaarten voorspelden zowel emotionele betrokkenheid als gedragsmatige betrokkenheid boven en boven eerdere niveaus van betrokkenheid.

De indirecte effecten van rapportcijfers via positief affect op zowel emotionele betrokkenheid als gedragsagressie waren significant. Hoe hoger de rapportcijfers, hoe lager het negatieve affect, wat vervolgens een hogere emotionele betrokkenheid voorspelde, maar geen gedragsmatige betrokkenheid.

Het indirecte effect van cijfers via negatief affect op emotionele betrokkenheid was sterker naarmate de leerlingen de prestatienormen in hun klas als hoog beschouwden. Dit indirect effect van negatief affect was ook sterker voor jongens dan voor meisjes, wat aantoont dat jongens met sterker negatief affect reageerden op hun cijfers dan meisjes.

Discussion

We toonden aan dat hogere cijfers na verloop van tijd een grotere emotionele en gedragsmatige betrokkenheid voorspelden. Tus, de resultaten van deze studie suggereren dat het belangrijk is om betrokkenheid te beschouwen als een reactie op prestatie feedback, niet alleen als een antecedent van schoolprestaties.

Leerlingen kunnen passief worden in het klaslokaal, niet zozeer omdat ze sterke negatieve gevoelens ervaren, maar eerder omdat ze minder positieve gevoelens ervaren.

Leerlingen zullen hun cijfers waarschijnlijk vergelijken met die van hun klasgenoten. Adolescenten die het gevoel hebben dat de normen voor prestaties in de klas hoog zijn, kunnen meer belang gaan hechten aan hun eigen prestaties, wat resulteert in een sterkere negatieve affectieve reactiviteit op hun rapport.

De negatieve affectieve reactiviteit van jongens op cijfers was sterker dan die van meisjes, wat resulteerde in sterkere effecten van cijfers op de emotionele betrokkenheid van jongens via negatief affect.

Cijfers stimuleren sociale vergelijkingen en bepalen de normatieve status van leerlingen in de peer group. Het kan de competitieve aard van cijfers zijn waardoor jongens er meer op reageren.

Conclusion



Deze studie suggereert dat lage cijfers een neerwaartse spiraal in gang kunnen zetten, waarbij opeenvolgende dalingen van de betrokkenheid resulteren in nog lagere cijfers. laag presterende leerlingen die zien dat hun medeleerlingen hoge cijfers halen, zijn bijzonder kwetsbaar. Jongens zijn kwetsbaar voor afnemende betrokkenheid, omdat zij meestal lagere cijfers krijgen en affectiever reageren op cijfers dan meisjes.

Docenten moeten zich bewust zijn van de mogelijke negatieve gevolgen van de cijfers die zij geven en proberen die te beperken.

1.2 Black & William (2018) Classroom assessment and pedagogy

Introduction

Er is een gebrek aan een adequate theorie van formatieve beoordelingen, deze moet worden ingebed in het bredere theoretische veld. Formatieve evaluatie kan niet volledig worden begrepen tenzij binnen de context van een pedagogische theorie.

Elke benadering van de verbetering van de praktijk in de klas, die gericht is op beoordeling, moet alle aspecten van beoordeling op een geïntegreerde manier behandelen. Het is daarom, denken wij, moeilijk te zien hoe formatieve evaluatie verder kan worden ontwikkeld zonder de relatie tussen formatieve en summatieve aspecten van het werk van leraren te verduidelijken.

Een beoordelingsactiviteit waarbij de leerlingen iets leren, is, als alle andere omstandigheden gelijk zijn, te verkiezen boven een activiteit waarbij zij dat niet doen. "Assessment as learning" is in een meer specifieke betekenis gebruikt: om de rol van leerlingen bij het controleren en sturen van hun eigen leren te omschrijven. De term "assessment of learning" wordt dan gebruikt om het proces te beschrijven waarbij leraren beoordelingsgegevens gebruiken om hun onderwijs te onderbouwen, en "assessment of learning" verwijst naar het gebruik van beoordeling om vast te stellen in hoeverre de leerlingen de vooropgestelde leerresultaten hebben bereikt.

Summatief vs. Formatief

Wanneer de conclusies betrekking hebben op de status van de leerling, of op zijn toekomstig potentieel, dan functioneert de beoordeling summatief. Wanneer de conclusies betrekking hebben op het soort acties dat de leerling het best zou helpen leren, dan functioneert de beoordeling formatief.

Dwars door het onderscheid tussen summatieve en formatieve functies van beoordeling zijn kwesties over wie beslist over de beoordeling, waar de beoordeling plaatsvindt, en hoe het werk van de leerlingen wordt beoordeeld.

De focus van dit artikel is dan ook om zowel formatieve als summatieve aspecten van klassikale beoordeling in te bedden in een breder theoretisch kader, zoals aanbevolen door Perrenoud, met behoud van een duidelijke focus op klassikale beoordeling als beoordeling - een procedure voor het maken van gevolgtrekkingen.

Pedagogy and/or instruction

Voor de doeleinden van dit document, definiëren we daarom pedagogie in het algemeen, en specifiek als "de handeling en het discours van het onderwijzen".

Pedagogie is de handeling van het lesgeven samen met het bijbehorende discours van onderwijstheorieën, waarden, bewijzen en rechtvaardigingen. Het is wat men moet weten, en de vaardigheden die men moet beheersen, om de vele verschillende soorten beslissingen te nemen en te rechtvaardigen waaruit het onderwijs is opgebouwd. Het leerplan is slechts één van zijn domeinen, zij het een centraal domein.

Dit gebrek aan aandacht voor beoordeling proberen wij aan te pakken. De nadruk ligt op de context van onderwijzen en leren in de klas.

Models for pedagogy and instruction

"Vier fundamentele vragen die beantwoord moeten worden bij het ontwikkelen van een curriculum en plan van instructie":

1. Welke onderwijsdoelstellingen moet de school nastreven?
2. Welke onderwijservaringen kunnen worden aangeboden die waarschijnlijk deze doelen te bereiken?
3. Hoe kunnen deze educatieve ervaringen doeltreffend worden georganiseerd?
4. Hoe kunnen we bepalen of deze doelen worden bereikt?

Model for assessment in relation to pedagogy:

Components of the model

The broad determinants of pedagogy and instruction

Het doel van de pedagogie: veel verklaarde doelen houden verplichtingen in over wat de leerlingen moeten leren, en ook over hoe zij dat moeten leren.

"Zij" zijn het erover eens dat de gekozen inhoudelijke thema's centraal moeten staan in de betrokken vakken en dat een hoofddoel moet zijn het vermogen van de leerlingen te ontwikkelen om wat zij in de ene context hebben geleerd, in een andere context toe te passen

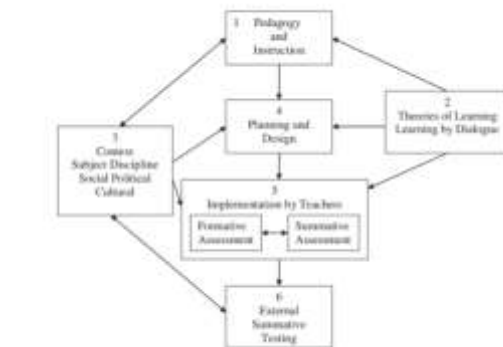


Figure 1. Model for assessment in relation to pedagogy

Planning and design

Planning moet op verschillende niveaus plaatsvinden, van algemeen en algemeen tot gedetailleerd en specifiek. Goede activiteiten kunnen de belangstelling wekken en ontwikkelen.

Implementation – teachers' formative assessment

De bekende bevindingen over de effecten van feedback benadrukken het belang van één-op-één-uitwisselingen van de leerkracht met elke leerling, hetzij in de vorm van schriftelijke (maar alleen van commentaar voorziene) feedback, hetzij mondeling. Een alternatief gebruik is het opzetten van peer-assessment groepen.

Het belangrijkste punt hier is dat de contingente aard van het leerproces - of het nu in dialoog is, wanneer een leraar commentaar schrijft op het werk van een leerling, of wanneer een leerling feedback krijgt van een medeleerling - betekent dat het verzamelen, identificeren en interpreteren van bewijsmateriaal een onmisbaar onderdeel is van effectieve instructie. Beoordeling als beoordeling is de kern van effectieve instructie.

Implementation – teachers' summative assessment

De waarde van evaluatieve feedback voor leraren om hen te helpen na te denken over hun werk.

Een uitzondering op het formatieve gebruik van summatieve toetsen is wanneer summatieve toetsen door de leraar werden gebruikt om feedback te genereren op dezelfde manier als ander schriftelijk werk. Deze manier helpt de leerlingen bij hun zelfreflectie en de feedback was eerder bedoeld om hun prestaties te verbeteren dan om een oordeel of een cijfer te geven.



- De eerste is dat toetsen ontworpen voor summatieve doeleinden een rol moeten spelen bij het verduidelijken van wat domeinbeschrijvingen eigenlijk betekenen.
- De tweede is dat toetsen kunnen worden gebruikt als leeractiviteiten.

Summative assessment by teachers

Er is een toenemende bezorgdheid dat de in scholen gebruikte beoordelingssystemen een breder scala van conclusies over de capaciteiten van leerlingen moeten ondersteunen dan tot nu toe het geval was.

Veranderingen in het soort conclusies dat beoordelingen geacht worden te ondersteunen, hebben op hun beurt problemen met traditionele tests en examens aan het licht gebracht, in het bijzonder het feit dat traditionele beoordelingen vaak geen zaken beoordelen waarover gebruikers van beoordelingsinformatie conclusies willen trekken.

Het is onmogelijk om te zorgen voor consistentie in de in een lokale onderwijsinstantie, laat staan in een heel rechtsgebied. In hoeverre dat wantrouwen gerechtvaardigd is, is opnieuw een punt van discussie, maar er lijkt weinig twijfel over te bestaan dat het typische niveau van wat men "beoordelingsgeletterdheid" (Stiggins, 1991) zou kunnen noemen, in de meeste landen vrij laag lijkt te zijn.

Wij moeten ons inderdaad bezighouden met de context van openbare, lokale of nationale beoordelingen, waar procedures en instrumenten het vertrouwen van het publiek moeten winnen. Maar er is een bredere kwestie omdat leraren en hun scholen regelmatig, en in veel gevallen vaak, summatieve beoordelingen van hun leerlingen moeten geven, vooral wanneer aan het eind van elke fase (b.v. elk schooljaar) beslissingen moeten worden genomen over cursussen en klassen voor de volgende fase.

The links between formative and summative assessment

Leerkrachten in staat zijn een bijdrage te leveren aan de summatieve doelstellingen van de beoordeling, zodat een systeem dat in de eerste plaats is ontworpen en wordt gebruikt om summatieve doeleinden te dienen, in feite ook leerbevorderend kan zijn. Het andere is de veronderstelling dat een scherpe discontinuïteit tussen formatief en summatief de norm is, te vervangen door de aanvaarding dat er synergie tussen beide zou moeten - en kunnen - bestaan.

Naarmate leerkrachten meer ervaring opdoen met de uitvoering van summatief werk op de manieren die nodig zijn om de kwaliteit van hun summatieve oordelen te verbeteren, zullen zij geconfronteerd worden met de noodzaak om betere beoordelingsinstrumenten te produceren, en dus het ontwerp van het leerwerk koppelen aan de kenmerken (a) en (b) hierboven, en dieper moeten nadenken over de geldigheid van hun oordelen, en dus de kenmerken (c) en (d) koppelen aan dit ontwerp en de rechtvaardiging ervan in het licht van de brede doelstellingen van hun pedagogie.

Formative and summative assessment in pedagogy

Een van de belangrijkste ideeën die aan dit document ten grondslag liggen, is dat beoordeling een essentieel aspect is van doeltreffend onderwijs, omdat de relatie tussen de instructie die leerlingen krijgen en wat zij als resultaat van die instructie leren, complex is. Vanuit deze eenvoudige visie zou er geen conflict moeten zijn tussen formatieve en summatieve beoordeling - sterker nog, het onderscheid zou niet nuttig zijn - omdat alle beoordeling zou gaan over het produceren van geldige gevolgtrekkingen over leerlingen. Beoordeling kan niet worden begrepen zonder rekening te houden met de ruimere context waarbinnen die beoordeling plaatsvindt.

1.3 Gerritsen et al. (2017) Assessment quality in tertiary education: An integrative literature review

Introduction

De kwaliteit van de beoordeling omvat de kwaliteit van alle aspecten van de beoordelingspraktijk. In het tertiair onderwijs hebben de gevolgen van een lagere beoordelingskwaliteit grote gevolgen voor de studenten en hun functioneren in de samenleving.

Consequences of inferior assessment quality

Een gebrekkige beoordeling kan het leren belemmeren, waardoor de prestaties kunnen dalen en het onderwijs helemaal niet kan worden afgemaakt. Een lagere kwaliteit van de beoordeling heeft ook gevolgen voor het selectie- en certificeringsdoel van de beoordeling, omdat het kan leiden tot oneerlijk zakken en oneerlijk slagen.

Bovendien heeft gebrekkige kwaliteit van de beoordeling gevolgen voor de verantwoordingsplicht van de beoordeling. Het kan leiden tot een onnauwkeurig beeld van de prestaties van leerlingen.

Difficulties in guaranteeing assessment quality

Het is belangrijk te verduidelijken hoe beoordelingskwaliteit wordt geconceptualiseerd en waarom er verschillen in conceptualisering bestaan. Een overzicht van deze perspectieven zal meer inzicht verschaffen in de betekenis van beoordelingskwaliteit.

Historical development of assessment

Bij de beoordeling lag de nadruk vooral op efficiëntie. In overeenstemming met het behaviorisme was observeerbaar gedrag belangrijker. Alleen gestandaardiseerde tests konden de vergelijking legitimeren.

Maatschappelijke eisen over wat studenten na hun afstuderen zouden moeten kunnen, veranderen in de loop der tijd als gevolg van economische en sociaal-culturele ontwikkelingen. Van studenten in het tertiair onderwijs wordt verwacht dat zij in staat zijn te anticiperen op en zich aan te passen aan veranderingen in hun toekomstige werkomgeving van levenslang leren. Kennis, vaardigheden en attitudes, alsook de toepassing van deze zogenaamde competenties in verschillende authentieke situaties. De focus verschoof van gedrag hoe leerlingen leren naar cognitieve en sociaal constructivistische gezichtspunten. Studenten gaan meer interacties aan en leren actief. Beoordeling werd geacht inzicht te verschaffen in de positie van leerlingen in hun leerproces, welke leerdoelen bereikt moesten worden en hoe deze doelen bereikt konden worden.

In de jaren zestig werd het onderscheid gemaakt tussen formatieve en summatieve evaluatie. Een evaluatie is formatief als zij wordt gebruikt voor de ontwikkeling of verbetering van een onderwijsproces. Een evaluatie is summatief als zij wordt gebruikt bij de besluitvorming over de eindresultaten van een onderwijsproces. Formatieve evaluatie levert informatie op die kan worden gebruikt om het onderwijzen en leren te veranderen, ook wel *assessment for learning* genoemd. *Assessment of learning* verwijst naar summatieve beoordeling.

Bovendien volstaat het niet langer de beoordeling uitsluitend vanuit psychometrisch oogpunt te beoordelen. Naast de traditionele kwaliteitscriteria, zoals objectiviteit, worden meer kwalitatieve criteria, zoals authenticiteit, educatieve consequenties en cognitieve complexiteit, gebruikt om de kwaliteit van beoordelingen te evalueren.

Samenvattend kan een lagere beoordelingskwaliteit nadelige gevolgen hebben voor het doel van het leren, voor de keuze en de evaluatie van de resultaten en voor de toerekenbaarheid. Deze studie heeft tot doel de evaluatiekwaliteit in het tertiair onderwijs te conceptualiseren.

Results

Validity

Validiteit is een beoordeling van de mate waarin het bewijsmateriaal en de argumenten de interpretatie en de afleidingen op basis van testcores ondersteunen. Validiteit verwijst naar het geaccumuleerde bewijsmateriaal over de doeltreffendheid van specifieke beoordelingsmodellen bij het meten van bekwaamheden. De interpretatie van validiteit is niet gebaseerd op een enkel statistisch gegeven, maar op een combinatie van al het beschikbare bewijsmateriaal over de beoordeling.

Does the test do what it claims to do? Does a test measure what it purports to measure?

Transparency

Transparantie wordt genoemd in verband met beoordeling vanwege de gevolgen ervan voor onderwijzen en leren. Transparantie wordt belangrijk geacht omdat inzicht in de basis waarop de prestaties zullen worden beoordeeld, de verbetering van de prestaties vergemakkelijkt. Transparantie is van essentieel belang omdat misverstanden en verkeerde interpretaties worden vermeden over de vraag welke prestaties worden beoordeeld en hoe dat gebeurt.

Reliability

Betrouwbaarheid wordt beschouwd als een eigenschap van de testscore. Het is belangrijk omdat het de nauwkeurigheid bepaalt van beslissingen om te slagen of te zakken. Betrouwbaarheid verwijst naar de mate waarin testcores consistent, precies en stabiel zijn, dus vrij van meetfouten. Elke variabiliteit (inconsistentie) wordt behandeld als een foutvariantie, en dus als een meetfout. Elk construct dat wordt gemeten is stabiel of wordt als stabiel beschouwd. De nadruk ligt op objectiviteit en standaardisering.

Betrouwbaarheid geeft de mate van genereerbaarheid van testcores aan; de score op de items in de steekproef moet indicatief zijn voor de scores van een leerling in elke andere relevante steekproef. Validiteit en betrouwbaarheid hangen met elkaar samen en kruisen elkaar. Betrouwbaarheid is een voorwaarde voor validiteit. Een test zal nooit geldig zijn als hij niet betrouwbaar is.

Influences on assessment quality criteria

Wat zijn de invloeden op de kwaliteit van de beoordeling? Vier thema's: standaardisatie, belanghebbenden, duidelijkheid en construct irrelevante variantie.

Standardisation

Standaardisatie is het eerste thema van invloeden. Vooraf bepaalde criteria en normen kunnen worden gebruikt om te beslissen of de test meet wat hij beweert te meten, en of de items en taken in de beoordeling representatief zijn voor de vereiste moeilijkheidsgraad. Het beoordelingsproces wordt transparanter wanneer alle elementen van een beoordelingsprogramma gedocumenteerd zijn. Standaardisering beïnvloedt de eerlijkheid van de beoordeling, wat een aspect van transparantie is. Standaardisatie beïnvloedt ook de betrouwbaarheid, meer consistentie en dus meer betrouwbaarheid.

Stakeholders

Belanghebbenden is het tweede thema van invloeden. Zij beïnvloeden de kwaliteit van de beoordeling door hun gebruik van de beoordeling, hun aanvaarding ervan of hun weerstand ertegen, de mate waarin zij de beoordeling belangrijk vinden, hun aandacht voor de beoordeling in plaats van voor zichzelf, hun betrokkenheid bij het ontwerpen van beoordelingsprogramma's en hun deskundigheid in de verschillende aspecten van het beoordelingsproces. De belanghebbenden beïnvloeden de validiteit door de manier waarop zij de beoordeling gebruiken. De beoordelingsscore weerspiegelt misschien niet volledig hun werkelijke prestaties, wat op zijn beurt de geldigheid ervan ondermijnt.



De perceptie van studenten van eerlijkheid, een aspect van transparantie, kan worden beïnvloed door de consistentie van de beoordelingsprocedures, de nauwkeurigheid van de informatie, teleurstellende beoordelingen en de aanwezigheid van vooroordelen.

Betrouwbaarheid kan worden beïnvloed door oorzaken van fouten, zoals lagere concentratieniveaus, enz. De betrouwbaarheid wordt vergroot door meer dan één beoordelaar te gebruiken om verschillende onderdelen van een beoordeling te beoordelen, zodat de bias die elke beoordelaar aan de beoordeling toevoegt, wordt beperkt.

Clarity

Duidelijkheid is het derde thema van de beïnvloeders over de kwaliteitscriteria voor beoordelingen. Een duidelijke beschrijving van de doelstellingen van een beoordelingsprogramma is van essentieel belang voor de kwaliteit van de beoordeling. Als leerlingen door onduidelijke taken hun ware niveau niet kunnen halen, neemt de validiteit af. Om de transparantie van het beoordelingsprogramma te garanderen, moeten de doelstellingen, regels en procedures van het programma, de rol en de rechten van de verschillende betrokkenen en de situaties waarin het programma moet worden gebruikt, allemaal duidelijk worden gedocumenteerd.

Individuele beoordelingen worden transparanter en consistentere als de betrokkenen weten wat er beoordeeld wordt. Wanneer de beoordelingscriteria of -onderdelen niet duidelijk of begrijpelijk zijn, kan de betrouwbaarheid afnemen.

Construct irrelevant variance

Construct irrelevante variance is de vierde invloed. Deze doet zich voor wanneer een assessment meer meet dan het geacht wordt te meten.

Evaluation of assessment quality criteria

Validation and Statistical data analyses

Validation

Validatie is het eerste thema dat verband houdt met de evaluatie van de criteria voor de kwaliteit van de beoordeling. Validatie wordt gebruikt om de validiteit te evalueren; zij wordt niet genoemd in verband met de betrouwbaarheid van de doorzichtigheid. Validatie evalueert of de interpretaties correct zijn, en dit is belangrijk omdat het de beslissingen op basis van testcores rechtvaardigt.

Statistical data analyses

Statistische gegevensanalyses zijn de tweede thema die verband houdt met de evaluatie van de kwaliteitscriteria voor de beoordeling. Analyses van de testonderdelen zijn nuttig omdat zij informatie verschaffen over de kenmerken van de testonderdelen, zoals de moeilijkheidsgraad, de discriminierendheid, de standaarddeviatie en de frequentie van afleiders.

De correlatie tussen testcores en maatstaven is een van de bewijsmiddelen die worden gebruikt om de validiteit van een test aan te tonen. Verschillende methoden kunnen de betrouwbaarheid van een test bepalen op basis van hun karakteristieken, nauwkeurigheid en bruikbaarheid voor een specifieke situatie: rest-hertest methode, splitt-half methode, spearman-brown formule.

Perspectives in the evaluation of assessment quality

Welke perspectieven wordt vastgesteld bij de evaluatie van de evaluatiekwaliteit? Er zijn vier thema's: Studenten, personeel, overheid en deskundigen.

De belanghebbenden stellen hun eigen kwaliteitscriteria vast of gebruiken hun eigen terminologie. Het perspectief van belanghebbenden kan een beoordelaar informatie geven over beoordelingskwaliteit. Het in aanmerking nemen van perspectieven bij het ontwerpen van beoordelingsprogramma's lijkt nuttig voor de aanvaarding van de beoordeling, voor de



betrokkenheid bij de beoordeling, voor het uitwerken van creatieve suggesties en voor het vaststellen van de geschiktheid voor de praktijk, die nodig is om kwaliteit te bereiken.

Students

Leerlingen gebruiken de beoordeling, dus ervaren ze de gevolgen ervan. Het is relevant om hun opvattingen over de beoordeling mee te nemen in de evaluatie. Vanuit het standpunt van de studenten zijn de componenten die de eerlijkheid van een beoordeling verhogen: relevantie van de taak voor het echte leven, redelijke eisen, actieve betrokkenheid van de student, ontwikkeling van een reeks vaardigheden, waardering van studietijd en inspanning door de beoordelingscore, en positieve langetermijneffecten.

Staff

De managers leken meer gericht op kwaliteitswaarborging, en dus op het voorkomen van fouten, dan op het vaststellen van de grote lijnen van een gemeenschappelijke visie op empirisch onderbouwde beoordeling.

Government

De relevantie van het perspectief van de overheid heeft vooral te maken met regels en voorschriften in verband met de kwaliteit van de beoordeling en de gevolgen daarvan voor andere belanghebbenden.

Experts

Deskundigen zijn materiedeskundigen of deskundigen die verbonden zijn aan een externe instantie. Deskundigen moeten worden betrokken bij de opstelling, herziening en uitvoering van beoordelingen.

Discussion

1. The themes of assessment quality criteria: validity, reliability and transparency
2. The themes of influences on assessment quality criteria: standardisation, stakeholders, clarity, and construct irrelevant variance.
3. The themes associated with the evaluation of assessment quality criteria: validation and statistical data analyses
4. The themes associated with the perspectives that should be acknowledged in the evaluation of assessment quality: students, staff, government and experts

2.1 Howard et al. (2018) What are standardised literacy and numeracy test testing? Evidence of the domain general contributions to students' standardised educational test performance.

Met deze studie kunnen beleidsmakers op het gebied van onderwijs bepalen in welke mate de vooropgestelde onderwijsnormen worden nageleefd en kunnen zij steun verlenen aan districten, scholen en individuen om beter aan deze normen te voldoen.

In een poging om domein-specifieke geletterdheid en rekenvaardigheid te beoordelen, kan een gestandaardiseerde beoordeling ook domein-gerelateerde cognitieve capaciteiten beoordelen. In dergelijke gevallen is het onduidelijk of de resultaten van de leerlingen overeenkomen met hun werkelijke lees-, schrijf- en rekenvaardigheid, dan wel of hun testcores worden beperkt door beperkingen in hun domeincongruente capaciteiten.



Grotere domein-general cognitieve eisen kunnen gepaard gaan met toetsen, en kunnen het vermogen van leerlingen om hun ontlukende academische vaardigheden te tonen, beperken. Als de testcores van de leerlingen worden beperkt door niet-doelgerichte vaardigheden, kunnen de onderwijssteun en onderwijstrajecten verkeerd worden gericht.

Working memory, non-verbal reasoning and standardised test performance

Een van de meest gebruikte domein-general cognitieve vaardigheden in onderwijsevaluaties is het werkgeheugen. De relatie tussen het werkgeheugen van leerlingen in het basisonderwijs en hun prestaties op de Scholastic Aptitude Test was consistent hoog voor alle domeinen. Het is daarom te verwachten dat het werkgeheugen geassocieerd is met testprestaties, gezien de rol die het speelt bij leren. Een dergelijk verband zou geen kwestie van testonzuiverheid zijn, maar eerder een artefact van het verschil in academisch presteren tussen leerlingen met een hoog en leerlingen met een laag WG. Niet-verbaal redeneren blijkt een sterke voorspeller te zijn van nationale standaardtoetsresultaten.

Hoewel het niet mogelijk of wenselijk is om een onderwijsevaluatie te maken die helemaal geen betrekking heeft op WG of redeneren, wordt een eenheidsmeetfout geïntroduceerd wanneer domein-general cognitieve vaardigheden een direct effect hebben op de testprestaties dat verder gaat dan wat nodig is om de beoogde vaardigheid aan te tonen.

Estimating domain-general contributions to domain-specific assessment

Het Rasch model transformeert ruwe gegevens in twee onafhankelijke parameters, bekend als (1) vaardigheidsschattingen en (2) item moeilijkheidsschattingen. De relatie tussen twee parameters wordt op een lineaire schaal uitgedrukt als een logistische functie van de relatieve afstand tussen de vaardigheid van een persoon en de moeilijkheid van het item. Het verschil tussen de ruwe score of de waargenomen gegevens en de theoretisch afgeleide Rasch-schattingen staan bekend als residuen. Deze residuen geven de variantie in performance weer die niet kan worden verklaard door het te meten construct en vertegenwoordigen dus dimensies/constructies buiten het te meten unidimensionale construct.

The current study

De uit Rasch afgeleide gestandaardiseerde persoons-item vaardigheidsschattingen en residuen die worden verkregen uit de gestandaardiseerde testprestaties van testnemers, kunnen een methode bieden om te onderzoeken in welke mate domein-generale processen van invloed zijn op gestandaardiseerde testprestaties en via welke weg zij deze invloed uitoefenen.

Result

Alles bij elkaar suggereren de resultaten dat de NAPLAN-resultaten van de leerlingen niet alleen een resultaat waren van de kerncompetenties die werden beoordeeld, maar waarschijnlijk ook van een reeks niet-gerichte vaardigheden. Onze resultaten suggereren dat dit effect (effect van WN op leren) werd bevorderd door non-verbaal redeneren in alle beoordeelde inhoudsgebieden.

Discussion

Indirecte effecten van het WG op academische prestaties worden verwacht, gezien de essentiële rol van het WG bij het leren. Zo worden ook directe verbanden verwacht tussen WG en studieprestaties. In de mate dat WG de testresiduen verklaart, zijn deze domein-algemene cognitieve vaardigheden echter een dimensie geworden die direct werd beoordeeld, maar die perifeer is ten opzichte van de doelvaardigheden.

Conclusion

De resultaten suggereren dat bij deze gestandaardiseerde beoordeling voor het hele land niet-beoogde vaardigheden werden verward met vaardigheden die wel het doel van de beoordeling waren. Dit lijkt zich niet alleen uit te strekken tot het effect van het werkgeheugen op het leren in meer genres, maar ook tot de eisen van verschillende beoordelingstypen en -

methoden. Dit heeft gevolgen voor het vermogen van leerlingen om echte bekwaamheid op beoordeelde gebieden aan te tonen en voor de onderwijsondersteuning en -voorzieningen die op basis van deze resultaten worden verstrekt.

2.2 Butler (2018) Multiple choice testing in education: Are the best practices for assessment also good for learning?

De alomtegenwoordigheid van meerkeuzetoetsen in het huidige onderwijs is het gevolg van het voordeel dat deze toetsen bieden ten opzichte van andere beoordelingsformaten. Meerkeuzetoetsen zijn relatief gemakkelijk te scoren, bieden een grotere objectiviteit bij de beoordeling en maken het mogelijk meer inhoud te behandelen door de tijd te verkorten die de deelnemers nodig hebben om de vragen te beantwoorden.

Sommige van deze adviezen gelden meer in het algemeen voor beoordelingen (vermijd strikvragen enz.), maar de meeste zijn specifiek voor het meerkeuzeformaat (maak elk mogelijk antwoord op een vraag ongeveer even lang enz.).

Hoewel toetsing vaak wordt gezien als een instrument voor beoordeling, is het ook een activiteit die tot leren leidt. Studies hebben aangetoond dat het maken van een meerkeuzetoets bevorderlijk is voor het leerproces.

Aangezien het gebruik van meerkeuzetoetsen zowel casuïstisch als beoordelend is, is het van belang na te gaan of de beste praktijken voor beoordeling overeenstemmen met de beste praktijken voor leren.

Om meerkeuzetoetsen doeltreffend als beoordelingsinstrument te gebruiken, is er behoefte aan discrimineerbaarheid, betrouwbaarheid en validiteit. Voor het leren moeten toetsen duurzaam en generaliseerbaar zijn.

Best practice #1: Avoid using complex item types or answering procedures

Uit de literatuur blijkt dat de kosten van complexe meerkeuzevragen veel hoger zijn dan de baten. Dezelfde conclusie kan worden getrokken uit onderzoek naar de invloed van complexe vraagtypen en antwoordprocedures op het leerproces. Het centrale probleem bij complexe itemtypes en antwoordprocedures is dat de grotere complexiteit vaak leidt tot grotere variabiliteit in de cognitieve processen die verschillende testdeelnemers toepassen bij het beantwoorden van hetzelfde item.

Best practice #2: Create items that require the engagement of specific cognitive processes

Het is van cruciaal belang voor de geldigheid van metingen dat het correct beantwoorden van een gegeven item vereist dat de testnemers een specifieke reeks gewenste cognitieve processen in gang zetten. Een hulpmiddel kan Bloom's taxonomie van leerdoelen zijn. Het is belangrijk na te gaan welke cognitieve processen zullen worden ingeschakeld bij het beantwoorden van het item. Het afnemen van meerkeuzetests die higher-order thinking vereisen, kan leiden tot een verbetering van de latere prestaties.

Best practise #3: Avoid using None of the above (NOTB) and All of the above (AOTB) as response options.

De beoordelingsliteratuur zegt het gebruik ervan te vermijden. NOTB-vragen ondermijnen het sleutelmechanisme dat het leren van toetsen ondersteunt (het terugvinden van de coreect informatie). Het gebruik van AOTA lijkt gunstig te zijn voor het aanleren wanneer het het coreect alternatief is. Wanneer AOTA het antwoord is, wordt de testnemer alleen blootgesteld aan correcte informatie, omdat alle alternatieven waar zijn.

Best practices #4: Use three plausible response options

Drie antwoordmogelijkheden bieden het beste evenwicht tussen psychometrische kwaliteit en efficiëntie van de uitvoering (leerlingen kunnen in dezelfde tijd meer drie alternatieve vragen beantwoorden dan vier alternatieve vragen). Het precieze aantal dat wordt gekozen, wordt het best bepaald door het aantal plausibele onjuiste antwoorden dat kan worden gecreëerd. Als efficiëntie geen punt van zorg is en er een plausibel foutief antwoord kan worden gegeven, dan is het gebruik van vier alternatieven aanvaardbaar.

Best practices #5: Create Multiple-Choice test that are challenging, but not too difficult

Het gebruik van meerkeuze toetsen voor het leren suggereert ook dat het het beste is om relatief weinig alternatieven te hebben, maar deze aanbeveling wordt gedaan om een andere reden. Een meerkeuzetoets die de proefpersonen blootstelt aan veel onjuiste informatie is zorgwekkend, omdat zij die mogelijkwijs kunnen leren. Negatief suggestie-effect.

Zoals verwacht kan worden, verhoogt het verhogen van het aantal antwoordmogelijkheden over het algemeen de moeilijkheidsgraad van het item, waardoor de kans toeneemt dat een testnemer een afleidend item incorreecteert. Een groter aantal alternatieven vermindert de positieve effecten en verhoogt de negatieve effecten op het leren; meer bepaald leren en onthouden de testnemers minder correcte informatie en verwerven en reproduceren zij ook meer incorrecte informatie.

Het juiste antwoord op een vraag gebruiken als lokmiddel voor een andere vraag is een gangbare praktijk in het onderwijs en een aanbevolen strategie om een effectieve afleider te creëren.

Wanneer leerlingen worden uitgedaagd, moet het leren worden gestimuleerd en moet er feedback worden gegeven.

Bonus best practice: Provide feedback

Feedback is een wezenlijk onderdeel van formatieve beoordeling. Feedback geven na een meerkeuzetoets versterkt de positieve effecten op het leren en vermindert of elimineert de negatieve effecten aanzienlijk.

2.3 Brookhart (2018) Appropriate criteria: Key to effective rubrics

De aanwezigheid van zowel criteria als beschrijvingen van het prestatieniveau onderscheidt rubric van andere soorten evaluatie-instrumenten.

Introduction

Een rubriek formuleert de verwachtingen voor het werk van de leerlingen door criteria voor het werk en prestatiebeschrijvingen op te sommen over een continuüm van kwaliteit. Een rubriek bestaat dus uit twee delen: criteria en beschrijvingen van de prestatieniveaus.

Een belangrijk kenmerk van een rubriek is of ze algemeen of taakspecifiek is. Taakspecifieke rubrics specificeren de specifieke feiten, concepten en/of procedures die de antwoorden van de leerlingen op de taken moeten bevatten. Rubrics helpen de leerlingen door de verwachtingen van de leerlingen expliciet te maken en, meer in het algemeen, te bepalen hoe leren eruit ziet. Rubrics spelen een rol in de formatieve leercyclus en ondersteunen student agency en zelfregulering.

Algemene criteria moeten in beschrijvende taal worden geschreven.



Een andere belangrijke manier om beoordelingscriteria te karakteriseren is of ze analytisch of holistisch zijn. Analytische rubrics beschouwen criteria één voor één (beter voor feedback). Holistische criteria beschouwen alle criteria tegelijk en vereisen slechts één beslissing over één criterium (beter om te beoordelen).

Rubrics worden gekarakteriseerd door het aantal criteria en het aantal niveaus. Het aantal criteria moet worden gekoppeld aan de beoogde leerresultaten die moeten worden beoordeeld, en het aantal niveaus moet worden gerelateerd aan de soorten beslissingen die moeten worden genomen en aan het aantal betrouwbare verschillen in het werk van de leerling dat mogelijk en nuttig is.

Formatieve evaluatie en het feit dat formatieve evaluatie wordt afgestemd op de verwachtingen van de leerlingen. De evaluaties helpen het leren van de leerlingen te verbeteren via verschillende mechanismen: ze verhogen de transparantie, verminderen de angst, helpen bij het feedbackproces, bevorderen de zelfeffectiviteit van de leerlingen of ondersteunen de zelfregulering van de leerlingen.

Purpose of this study

De beschrijving van de criteria en de prestatieniveaus samen kunnen de leerlingen helpen hun leerobjectieven te conceptualiseren, zich te concentreren op belangrijke aspecten van leren en presteren, en zich een beeld te vormen van waar zij staan in hun leerproces en wat zij moeten proberen te verbeteren.

This is a literature review

Results and discussion

Relation of rubric type to reliability, validity and learning

Betrouwbaarheid werd het meest bestudeerd als interbeoordelaarsbetrouwbaarheid. Constructvaliditeit werd bestudeerd in de vorm van expert review tot factoranalyses. Er werden geen waarneembare patronen gevonden die erop wezen dat de ene vorm van rubric de voorkeur verdiende boven de andere wat betrouwbaarheid of geldigheid betreft.

Conclusie

Rubrieken worden steeds duidelijker als onderdeel van de beoordeling in het hoger onderwijs. Drie, vier of vijf zijn de meest voorkomende als niveaus, de meeste van de rubrics zijn beschrijvend. Slechts 56% van de studies maakt melding van het gebruik van rubrics bij studenten.

In verschillende artikelen wordt uitdrukkelijk gesteld dat het doel van de rubrics is de verwachtingen van de opdrachten expliciet te maken.

Passende criteria zijn de sleutel tot effectieve rubrics. Als het criterium substantieel is, wordt de aandacht voor kwaliteit een deel van het werk.

6.1 Van der Gaag et al. (2019) Evidence for measuring teachers' core practices

Introduction

Om een hoog onderwijsniveau te bereiken dat het leren van de leerlingen positief beïnvloedt, is het van cruciaal belang feedback van hoge kwaliteit te geven over de praktijken van de leerkrachten. Kernpraktijken; de belangrijkste professionele activiteiten die leerkrachten moeten uitvoeren, onderstrepen het belang van de kennis van leerkrachten in actie. Kernpraktijken zijn bedoeld om aan te sluiten bij de dagelijkse praktijk van leerkrachten en om inhoudelijk hoogwaardig, zinvol onderwijs te ondersteunen. De kernactiviteiten van leerkrachten worden vaak ingedeeld in (1) activiteiten vóór de les, (2) lesactiviteiten en (3) activiteiten na de les.

Er is weinig bekend over het bewijs dat wordt geleverd voor het gebruik van meetinstrumenten in termen van validiteit, betrouwbaarheid, bruikbaarheid en het effect dat zij kunnen hebben op de ontwikkeling van leraren. Deze reviewstudie geeft een overzicht van de evidentie die uit studies naar voren komt over de metingen van de kernpraktijken van leraren.

Validity

Bewijs voor de geldigheid houdt in dat de relatie tussen de kernpraktijken van de leraren, de doelstellingen en de gevolgen van een meting wordt bewezen. Bewijs voor de validiteit is nodig om systematische meetfouten te beperken en impliceert (1) inhoudsvaliditeit, (2) constructvaliditeit en (3) criteriumvaliditeit.

Reliability

Het bewijs voor de betrouwbaarheid heeft betrekking op de consistentie van de meting om willekeurige meetfouten te beperken en impliceert (1) interne consistentie, (2) stabiliteit en (3) interbeoordelaarsovereenstemming.

Bruikbaarheid betekent dat het instrument transparant, haalbaar en doeltreffend is om op de werkplek te gebruiken. Impact verwijst naar de gemaakte interpretaties en de gevolgen van de feedback.

De onderzoeksvraag luidt: Welk bewijs wordt geleverd voor de kwaliteit van de meting van de kernpraktijken van leerkrachten?

Results

Measurement instruments and assessors

- Looking at the type of teachers' core practices
 - o Instruments are mostly aimed at measuring lesson-related activities
 - o Less interest in measuring post lesson activities
 - o Barely aimed at measuring the pre-lesson activities
- Different types of instruments
 - o Questionnaires
 - o Observation instruments
 - o Interviews
 - o Performance and logbooks
- Different kinds of assessors
 - o Teachers self-report
 - o Students' assessment of their teachers
 - o Perceptions of colleagues and supervisors.

Evidence for instrument quality

Uit de beoordeling van de kwaliteit van de validiteit bleek dat de meeste instrumenten als "medium evidence" werden beoordeeld. Analyse van het aspect betrouwbaarheid wees uit dat voor de meeste instrumenten ten minste enig bewijs werd geleverd. Uit de analyse van de

bruikbaarheidskwaliteit bleek dat de meeste instrumenten geen bewijs leverden. Met andere woorden, er werd geen informatie verstrekt over de transparantie, de uitvoerbaarheid en de doeltreffendheid van het instrument.

De analyse van het impactkwaliteitsaspect omvatte een theoretische onderbouwing van de formatieve of summatieve consequenties. Specifiek bewijs over hoe een instrument de professionele ontwikkeling van leerkrachten zou kunnen bevorderen, ontbrak.

Discussion

Uit de bevindingen bleek dat de instrumenten vooral gericht waren op het meten van de kernpraktijken van de leerkrachten met betrekking tot de lessen, met name het klimaat in de klas en de didactische activiteiten. Er werd geen aandacht besteed aan het meten van de professionaliseringsactiviteiten van leraren. Interventies in de praktijken van leerkrachten kunnen als nuttig worden beschouwd en vergen vaak rehabilitatie, feedback en samenwerking om te worden uitgevoerd.

De overkoepelende interpretatie van de verkregen bevindingen voor elk type instrument, ongeacht de beoordelaar, degelijke theoretisch en empirisch bewijs voor de kwaliteit ervan ontbreekt.

Zoals eerder aangegeven, zijn deze kwaliteitsaspecten bij de meeste meetinstrumenten ook voor verbetering vatbaar. Zonder een degelijke beschrijving van de kwaliteit van instrumenten om de kernpraktijken van leerkrachten te meten, moeten gebruikers voorzichtig zijn bij het interpreteren van de verkregen scores.

Over het geheel genomen is het belangrijkste doel van het meten van de kernpraktijken van leraren het geven van feedback aan de leraren, die de professionele ontwikkeling kan stimuleren. In dit opzicht is impact een van de belangrijkste kwaliteitscriteria.

Het is waardevol om informatie over hoe leerlingen leren te integreren in hoe leraren lesgeven. Nuttige voorbeelden zijn: multisource feedback, mogelijkheden voor peer review en peer feedback van en door leraren, het betrekken van leerlingen bij de evaluatie van het onderwijs, de rol van teamwerk in het onderwijs.

6.2 O'Connel et al. (2016) Does calibration reduce variability in the assessment of accounting learning outcomes?

Onze bevindingen suggereren dat consensuele matiging de variabiliteit tussen de beoordelaars vermindert en ook het vertrouwen van de beoordelaar doet toenemen.

Introduction

Een leernorm is het vaste prestatieniveau dat van een leerling wordt verwacht om een erkend cijfer te krijgen, terwijl een leerresultaat het leren is dat uit het feitelijke werk van de leerling blijkt. De wens om vergelijkbaarheid van instellingsoverstijgende leerresultaten te bereiken. Er is veel belangstelling voor de vraag hoe leerresultaten het best kunnen worden beoordeeld. De betrouwbaarheid en validiteit van zowel de interne borging van leerresultaten als de daaropvolgende externe beoordeling zijn afhankelijk van het feit of interne en externe beoordelaars vergelijkbare oordelen vellen. Gelijksortige oordelen zijn onwaarschijnlijk, aangezien elke veronderstelling dat de oordelen van de beoordelaars, met name over niet-kwantitatieve leerresultaten, consistent zijn, in twijfel kan worden getrokken, aangezien zij slecht onderbouwd, grillig en onbetrouwbaar kunnen zijn.

Verscheidene onderzoekers hebben twijfels geuit over de vraag of examinatoren dergelijke normen wel consistent kunnen beoordelen.

De beoordelaars nemen deel aan consensusmoderatie process, wat na verloop van tijd zal resulteren in een vermindering van de variabiliteit van de beoordelaars. In deze studie wordt

nagegaan hoe groot het probleem van de variabiliteit van de beoordelaars is en of een consensusmoderatieproces de variabiliteit van de oordelen van de beoordelaars kan verminderen.

Er is weinig onderzoek gedaan naar de betrouwbaarheid van de beoordelaar op basis van een rigoureuze experimentele opzet.

De focus van onze studie ligt op de vraag of een kalibratie workshop waar beoordelaars samenkomen om te discussiëren en normen toe te passen op het werk van studenten, consensus kan bereiken in het toepassen van drempel leerstandaarden op het werk van studenten.

Om de omvang van het probleem van de variabiliteit van de beoordelaars te begrijpen, ontwikkelen we een duurzaam model om te garanderen dat de drempelnormen voor het leren worden gehaald door externe peer-beoordelaars te gebruiken.

Een grotere consistentie in de beoordelingen van de beoordelaars bereiken door een rigoureuze ijkproces te ontwikkelen voor beoordelingen van het leren op basis van drempelnormen. Het kalibratieproces is een interventie in verschillende fasen waarbij externe collega-beoordelaars zowel de beoordelingstaak als de voorbeelden van het werk van de studenten toetsen aan de drempelleernormen vóór, tijdens en na een kalibratieworkshop.

The context of the learning and teaching environment

De toenemende concurrentie heeft geleid tot een roep om meer transparantie en verantwoording van onderwijsstelsels en -resultaten, met inbegrip van een gezamenlijke inspanning om ervoor te zorgen dat de leerresultaten van afgestudeerden voldoen aan prestatienormen, zonder dat daarbij wordt teruggegrepen op gestandaardiseerde resultaten of tests.

Deze ontwikkelingen hebben geleid tot meer aandacht voor de waarborging van de kwaliteit van academisch onderwijs in het algemeen, met inbegrip van de integriteit en de degelijkheid van de beoordelingspraktijken van de verstrekkers van hoger onderwijs.

Hoewel de drempelnormen voor boekhouden op nationaal niveau zijn goedgekeurd, is er nog geen gedeelde opvatting over de wijze waarop deze normen eerlijk, consequent en kosteneffectief moeten worden beoordeeld.

Leerstandaarden van een hogere orde, zoals toepassingsvaardigheden en inzicht, zijn niet rechtstreeks meetbaar en vatbaar voor individuele interpretatie.

Review of literature on developing a shared understanding of academic standards

Het beoordelingsproces van students work is niet consequent en niet eerlijk. Een betrouwbaar beoordelingsproces levert consistente beoordelingsresultaten op. Onderzoekers hebben echter gewezen op een aanzienlijke variabiliteit tussen de beoordelingscriteria van de beoordelaars en de toegekende cijfers.

Academici kennen kwalificaties toe aan het werk van studenten op grond van hun uiteenlopende professionele kennis en ervaring. Cijfers verschillen in de wijze waarop ze worden verdeeld; de betrouwbaarheid van de beoordelaars is lager bij examens in de vorm van opgaven en problemen, en omgekeerd hoger bij kennis die moet worden teruggeroepen.

Om de inherente problemen van variabiliteit bij de beoordeling aan te pakken, zijn pogingen ondernomen om de betrouwbaarheid en validiteit van de beoordeling te vergroten door het gebruik van beoordelingscriteria, cijferbeschrijvingen en rubrics. Meer "vage normen kunnen echter niet worden omgezet in scherpe normen door gewoon meer gedetailleerde taal te gebruiken".

Social constructivism

De hierboven aangehaalde literatuur steunt de stelling dat variatie in de leerresultaten en de beoordeling van normen problematisch is. Er zijn twee paradigma's die van toepassing zijn op de beoordeling van leerresultaten: Techno-rationalistisch en sociocultureel. De techno-rationalistische epistemologie sluit aan bij de positivistische opvatting dat normen kunnen worden geëxpliciteerd via een op criteria gebaseerde aanpak, en vervolgens op objectieve wijze

kunnen worden geïnterpreteerd. Het socioculturele paradigma daarentegen gaat ervan uit dat beoordeling een contextafhankelijke, sociaal-situationele, interpretatieve activiteit is.

De hierboven aangehaalde literatuur steunt de stelling dat variatie in de leerresultaten en de beoordeling van normen problematisch is. Er zijn twee paradigma's die van toepassing zijn op de beoordeling van leerresultaten: Techno-rationalistisch en sociocultureel. De techno-rationalistische epistemologie sluit aan bij de positivistische opvatting dat normen kunnen worden geëxpliciteerd via een op criteria gebaseerde aanpak, en vervolgens op objectieve wijze kunnen worden geïnterpreteerd. Het socioculturele paradigma daarentegen gaat ervan uit dat beoordeling een contextafhankelijke, sociaal-situationele, interpretatieve activiteit is.

Beoordelingsontwerp en -proces zijn contextueel en sociocultureel geconstrueerd.

De beginselen van de beoordeling, die academici in staat stellen voldoende gedeelde opvattingen over academische normen te ontwikkelen om consistente beoordelingsbeslissingen te nemen met voortdurende controle en herkalibratie, worden op soortgelijke wijze opnieuw geformuleerd met behulp van de term consensus moderatie.

Wij erkennen dat het geven van cijfers een complexe en inherent problematische bezigheid is, waarbij niet definitief kan worden verklaard dat een cijfer dat door de ene academicus wordt gegeven hetzelfde is als het cijfer dat door een andere academicus voor hetzelfde werk wordt gegeven. Consensusmoderatie is een proces om te garanderen dat een beoordelingsresultaat geldig, eerlijk en betrouwbaar is en dat de beoordelingscriteria consequent zijn toegepast.

Summary and conclusions

We meten de mate van variabiliteit van de grader of de grootte van het probleem. De waarschijnlijke implicatie van dergelijke gelijkheid in de beoordeling is dat werkgevers en andere belanghebbenden zich niet op deze beoordelingen kunnen verlaten als betrouwbare beoordelingen.

Als de beoordelingsvariatie kan worden verminderd, kunnen de studenten en andere belanghebbenden met meer vertrouwen op de beoordelingen van de beoordelaars van de capaciteiten van de studenten. Een oplossing voor dit probleem bestaat erin processen te ontwikkelen die een gemeenschappelijk inzicht bevorderen in de wijze waarop de resultaten moeten worden beoordeeld, en zo de beoordelingsverschillen tussen beoordelaars te verkleinen.

De behandeling resulteerde in een significant verminderde variabiliteit onder de beoordelaars en een toegenomen vertrouwen van de beoordelaars in hun vermogen om de beoordelings- en toepassingsvaardigheden van de studenten te beoordelen, wat het potentieel aantoont voor incest in het ontwikkelen van een gedeeld begrip van standaarden door middel van een workshop kalibratieproces.

6.3 Bok et al. (2018) Validity evidence for programmatic assessment in competency-based education

Competentiegericht onderwijs is nu alomtegenwoordig. Een basisprincipe van CBE is het beoordelen en identificeren van de progressie van competentieontwikkeling bij studenten.

Introduction

Een dringende kwestie waarmee het competentiegericht onderwijs wordt geconfronteerd, is het waarborgen van het haalbare gebruik van degelijke, aan de prestaties gerelateerde informatie PRI waarmee feedback over het leren kan worden gegeven en beslissingen kunnen worden genomen over de vooruitgang op het gebied van de leerstof. Programmatische beoordeling in CBE brengt de volgende kenmerken en complexiteiten met zich mee.



1. Het maakt de instructie, begeleiding, supervisie en evaluatie van meervoudige affectieve, gedragsmatige en cognitieve prestatiedimensies mogelijk.
2. Prestatiebeoordelingen bieden de mogelijkheid voor student-specifieke, prestatie-relevante informatie. PRI, die wordt gedefinieerd als de formeel verzamelde informatie, zowel kwantitatief als kwalitatief, die informatie geeft over het leren van studenten binnen een bepaalde context, is belangrijk omdat zij gegevens bevat over de beoordeling van de prestaties in de tijd, en dus over het leren. Deze performance-oordelen kunnen echter door meer dan alleen de RPI worden beïnvloed en daarom is het van vitaal belang de impact te begrijpen van variabelen zoals de beoordeelde competentie, de beoordelingsmethode en het eind van de prestaties.
3. Een verscheidenheid van prestatie-indicatoren moet dienen als input voor prestatiebeoordelingen.
4. Een programmatische aanpak van de beoordeling maakt het mogelijk de veranderingen binnen en tussen individuen in de loop van de tijd te begrijpen. Veranderingen in de prestaties binnen een individu geven inzicht in de specifieke groeipatronen van het individu in de loop van de tijd. Veranderingen tussen individuen geven inzicht in de variatie van individueel leren in de loop van de tijd.

Het ideale model voor programmatische evaluatie zou zowel voortdurende PRI moeten verschaffen betreffende de ontwikkeling van studenten binnen vooraf bepaalde competenties als CBE met curriculaire programmaresultaten moeten verenigen.

Er zijn vier referenties voor validiteit, waarvoor voldoende geldigheidsargumenten moeten worden aangevoerd, waaronder score, generalisatie, extrapolatie en implicaties.

Conclusion

Deze studie levert validiteitsbewijs met betrekking tot interferenties die gemaakt zijn over de klinische prestatiescores van studenten bij programmatische beoordeling. We pasten een beoordelingsprogramma toe in een competentiegericht klinisch curriculum dat gericht was op integrated leren en beoordelen, en het resultaat ondersteunt onze beweringen dat het de studenten helpt om hun learning te maximaliseren en tegelijkertijd robust summatieve besluitvorming mogelijk maakt. Leeranalyses helpen dus om een overzienbare synthese te maken van PRI om de ontwikkeling van competenties in de tijd te volgen, zowel voor formatieve, summatieve als kwaliteitscontrole doeleinden.

7.1 Boud & Soler (2016) Sustainable assessment revisited

Duurzame evaluatie is voorgesteld als een idee dat gericht is op de bijdrage van evaluatie aan het leren voorbij de tijdschaal van een cursus geven. Het werd geïdentificeerd als een beoordeling die beantwoordt aan de behoeften van het heden in termen van de eisen van formatieve en summatieve beoordeling, maar die de studenten ook voorbereidt om aan hun eigen toekomstige leerbehoeften te voldoen.

Introduction

Onderwijs moet worden beoordeeld naar de resultaten en de gevolgen ervan. Duurzaamheid in het onderwijs kan worden geïnterpreteerd als een kenmerk van onderwijsstelsels. Duurzaam onderwijs gaat over de vraag welke onderwijspraktijken nu nodig zijn om leerlingen te vormen en op te leiden die in staat zullen zijn effectief te functioneren in een complexe, voortdurend veranderende samenleving. Duurzaam leren is dus een functie van wat leerlingen uit het onderwijs halen, niet van de input die in het proces wordt gestopt.

Defining and elaborating sustainable assessment

Duurzame evaluatie werd ontwikkeld om de nadruk te leggen op de noodzaak dat alle evaluatiepraktijken de lerende toerusten voor de uitdagingen van het leren en de praktijk waarmee zij te maken zullen krijgen zodra hun huidige leerepisode is voltooid.

De opvatting van beoordeling als 'geïnformeerde beoordeling' is een verschuiving naar beoordeling die wederzijds wordt opgebouwd tussen leerlingen en beoordelaars/docenten. Een geïnformeerde beoordeling van de eigen capaciteiten, praktijken en verworvenheden is niet alleen iets dat leerlingen moeten ontwikkelen om doeltreffend te kunnen leren, maar is ook nodig voor anderen, zoals leraren, om een oordeel te kunnen vellen dat ofwel gebruikt kan worden om leerlingen te adviseren, ofwel formeel geregistreerd kan worden als een indicator van hun vorderingen of prestaties.

Het omvat het vermogen om bewijsmateriaal te evalueren, situaties en omstandigheden scherp te beoordelen, deugdelijke conclusies te trekken en in overeenstemming met deze analyse te handelen. De sleutelementen voor de ontwikkeling van een gefundeerde oordeelsvorming vanuit het perspectief van de studenten werden voorgesteld als:

1. Identifying oneself as an active learner
2. Identifying one's own level of knowledge and the gaps in this
3. Practising testing and judging
4. Developing these skills over time, and
5. Embodying reflexivity and commitment

De rol van feedback bij de ontwikkeling van het leervermogen van studenten is enthousiast ontvangen. In dit meer recente werk ligt de nadruk op de bijdrage van feedback aan het leren door beoordeling, en op het herpositioneren van het begrip feedback, niet als een daad van informatieverstrekking aan studenten, maar als een coproductief proces waarin zowel de studenten als anderen een sleutelrol te spelen hebben. Leren kan in geen enkel opzicht duurzaam zijn als het van de leraren voortdurende informatie over het werk van de leerlingen vergt.

How has sustainable assessment been taken up?

Wat de ontwikkeling van een duurzame evaluatie betreft, zijn twee hoofdrichtingen in de literatuur duidelijk. De eerste is het toepassen van de ideeën op specifieke situaties of bepaalde contexten. De tweede richting is het ontwikkelen van bepaalde praktijken die worden besproken als onderdeel van duurzame beoordeling.

1. Applications in particular contexts

Een belangrijke strategie voor de toepassing van beoordelingspraktijken is ICT. Het gebruik van contextgebaseerde taken stelt leerlingen in staat zich te ontwikkelen tot effectieve levenslange beoordelaars.

2. Development of particular practices

- Self-assessment

Duurzame evaluatie impliceert de ontwikkeling van zelfevaluatie door middel van nieuwe evaluatie-instrumenten. Hierdoor kunnen studenten zich bewust worden van hun eigen leerbehoeften en kan de docent hen de nodige vaardigheden aanreiken om te blijven leren.

- Use of peers

- Reflection and the use of portfolios

Reflectie-activiteiten met verschillende vormen van peer learning bieden leerlingen ervaring met zelf-begeleiding en scheppen zo een oordeel over hun eigen leerproces en dat van anderen.

- Positioning assessment as part of learning activities

What does an emphasis on sustainable assessment contribute to assessment practice?



Willen beoordelingstaken als duurzaam worden beschouwd, dan moet het hele beoordelingsproces worden opgevat als een actief onderdeel van het leerplan om studenten in staat te stellen bepaalde resultaten te bereiken, en niet alleen als een middel om na te gaan of resultaten al dan niet zijn bereikt. Dit betekent dat beoordeling bewust en holistisch moet worden opgezet om de leerdoelen te bevorderen, met inbegrip van het beheer van het leerproces door de studenten, en in de loop van een cursus moet leiden tot activiteiten die het mogelijk maken te demonstreren wat geleerd is.

Directions for sustainable assessment

Het doel van een duurzame beoordeling, namelijk de studenten uit te rusten voor het leren na hun cursus, is het foudnat voor ontwikkeling.

Design Features

Alle aspecten:

- Purpose
- Assessment tasks
- Dispositions and engagement
- How students are to be judged?

Alle hierboven besproken aspecten moeten bij elkaar worden gebracht door het ontwerp van de cursus, met name door het ontwerp van evenementen en activiteiten die voorafgaan aan, gepaard gaan met en volgen op de beoordelingstaken. Toetsingstaken staan niet op zichzelf; zij maken altijd deel uit van een opeenvolging van activiteiten, ofwel gespecificeerd door cursusvereisten, ofwel voorgesteld of uitgevoerd door docenten. Het gebruik van duurzame evaluatie is een manier om evaluatie te verweven met onderwijzen en leren. Het kan een vorm van constructieve afstemming op lange termijn zijn om beoordeling en leren op langere termijn dichter bij elkaar te brengen.

Het is niet voldoende dat veel wenselijke kenmerken aanwezig zijn. De sociaal-culturele context van onderwijzen, leren en beoordelen kan nog steeds samenspannen om goede bedoelingen en blijkbaar ook een goed ontwerp te dwarsbomen.

Duurzaamheid bij de beoordeling moet een virtuele cirkel zijn: naarmate de studenten beter in staat zijn beslissingen over hun eigen leren te nemen, worden zij effectievere leerlingen, zoals blijkt uit de resultaten die via beoordeling worden beoordeeld. Hierdoor kan de beoordeling meer op duurzaamheid worden gericht in plaats van alleen maar op de huidige prestaties.

Conclusion

Het is uitgebreid tot de ontwikkeling van geïnformeerde oordeelsvorming, het gebruik van zelf- en collegiale toetsing en de ontwikkeling van zelfregulering. Het potentieel van duurzame evaluatie, samen met vele andere initiatieven op het gebied van formatieve evaluatie, moet echter nog volledig worden benut. De tijdschaal voor beoordelingsverandering is erg lang, dus snelle veranderingen zijn.

7.2 Panadero et al. (2017) Effects of self-assessment on self-regulated learning and self-efficacy: four meta-analysis

Deze meta-analytische review onderzoekt de effecten van zelfbeoordeling op het zelfregulerend leren en de zelfeffectiviteit van studenten.

Introduction

De relatie tussen het construct zelfbeoordeling, zelfregulerend leren en self-efficacy is het voorwerp geweest van empirisch onderzoek. Deze relatie is wederkerig en ingewikkeld: zelfevaluatie wordt geconceptualiseerd als een regulerende leerstrategie, zelfgereguleerd leren is afhankelijk van zelfevaluatie - via zelfmonitoring en zelfevaluatie - om het leren van studenten

te ondersteunen; en self-efficacy wordt geacht de activatie van studenten en het gebruik van regulerende strategieën, zoals monitoring en evaluatie, te versterken. Bovendien kan zelfevaluatie de waargenomen bekwaamheid van studenten verhogen, wat de zelfeffectiviteit van studenten kan beïnvloeden.

Self-assessment

Zelfevaluatie werd gedefinieerd als een grote verscheidenheid van mechanismen en technieken waarmee students de kwaliteiten van hun eigen leervoorkeuren en -knobbels beschrijven en er eventueel verdienste of worth aan toekennen. Zelfevaluatie gaat over leerlingen die hun eigen werk beoordelen.

Zelfevaluatie is een fundamenteel onderdeel van formatieve evaluatie omdat het uiteindelijk de leerling zelf is die de kloof tussen een huidige prestatie en de gewenste norm moet dichten.

Self-assessment effects on students learning and performance

Er zijn aanwijzingen dat zelfevaluatie kan leiden tot een verbetering van de doeltreffendheid en de kwaliteit van het leren. Naast rubrics, waarvan is gesuggereerd dat ze het leren van studenten ondersteunen als ze worden gecombineerd met zelfevaluatie of andere metacognitieve activiteiten, is ook gebleken dat feedback de relatie tussen zelfevaluatie en leren beïnvloedt.

Self-assessment effects on self-regulated learning and self-efficacy

De voordelen van zelfevaluatie worden in verband gebracht met de versterking van de eigen verantwoordelijkheid voor het leren en het gebruik van zelfregulerende leerstrategieën. Wat zelfeffectiviteit betreft, zijn er volgens de sociaal-cognitieve theorie vier factoren die de zelfeffectiviteit van een individu verhogen, namelijk eigen ervaring van succesvolle prestaties, het zien slagen van anderen, aanmoediging en fysiologische factoren. De basisveronderstelling waarom zelfevaluatie een effect heeft op de zelfeffectiviteit van studenten is dat studenten, door een beter begrip te krijgen van de eisen die aan de te maken taak worden gesteld, waarschijnlijk beter zullen presteren en daardoor succesvolle prestaties zullen ervaren. Dit, op zijn beurt, zou leiden tot een gevoel van inzet en een perceptie van verbeterde bekwaamheid, wat een grote invloed zal hebben op de mate van zelfeffectiviteit.

Self-regulated learning

Volgens een van de meest gebruikte definities is SRL zelfgegenereerde gedachten, gevoelens en handelingen die gepland en cyclisch aangepast zijn aan de verwezenlijking van persoonlijke doelstellingen. Het gebruik van SRL-strategieën was een belangrijke voorspeller van academische prestaties. Er bestaan verschillende modellen van SRL, maar de meeste omvatten een voorbereidende fase, een uitvoeringsfase en een evaluatiefase, die elk uit verschillende subprocessen bestaan.

Self-regulated learning effects on student learning and performance

Uit de SRL-acties blijkt dat zij doeltreffend zijn. De SRL-strategieën zijn onderverdeeld in grotere categorieën. De sterkste effecten op leren voor volwassenen waren de constructen doelniveau, volharding, inspanning en self-efficacy.

Self-regulated learning and self-assessment

Er is een sterk theoretisch verband gesuggereerd tussen SRL en zelfevaluatie. De SRL-subprocessen vertonen een duidelijke gelijkenis met zelfevaluatie, aangezien ze gericht zijn op de evaluatie van de eigen prestaties. Voorgesteld wordt dat zelfevaluatie de sleutel is tot de internalisering van standaarden, zodat de leerling zijn eigen leerproces doeltreffender kan regelen, wat de eerste fase van SRL inhoudt (voorbereidende fase).

Self-efficacy

Self-efficacy is de overtuiging over de persoonlijke capaciteiten om een taak uit te voeren en de gestelde doelen te bereiken. Zelfeffectiviteit blijkt de sterkste voorspeller te zijn van academische prestaties. De invloed van self-efficacy op de conceptualisering en ontwikkeling van SRL werd onderzocht. Zelfeffectiviteit is een van de belangrijkste variabelen geworden, niet alleen in het onderzoek naar motivatie, maar ook in het onderzoek naar SRL, en daarom werd zelfeffectiviteit opgenomen in SRL-modellen.

Moderating variables

Er zijn vier modererende factoren: gender, leeftijd, self-assessment intervention and implementation agent (docent of onderzoeker).

Research questions

1. Do self-assessment interventions have an effect on students' SRL?
 - a. Is there a differential effect based on different SRL instruments and constructs?
2. Do self-assessment interventions have an effect on students' self-efficacy?
3. Do the moderating variables gender, age/educational level, self-assessment interventions, and implementation agent influence the effects of self-assessment on students' SRL and/or self-efficacy?

Results

De resultaten ondersteunen het bestaan van een positief effect van zelfevaluatie op de zelfevaluatie. De effectgrootte is relatief klein.

Negatieve SRL staat voor zelfregulerende handelingen die geacht worden nadelig te zijn voor het leren. Lagere waarden in relatie tot deze variabele moeten worden geïnterpreteerd als positief voor het leren van de leerlingen. De resultaten ondersteunen het positieve effect van zelfbeoordeling op zelfregulerend leren, met een middelmatige effectgrootte.

De resultaten ondersteunen het positieve effect van zelfevaluatie op self-efficacy, met een grotere effectgrootte. De bevindingen suggereren dat zelfbeoordelingsinterventies een groter effect kunnen hebben op de self-efficacy van meisjes, in vergelijking met jongens. Een tweede bevinding suggereert dat monitoring een krachtige component is van self-assessment als het gaat om het bevorderen van de self-efficacy van leerlingen en dat het niet gebruiken van een self-assessment instrument de self-efficacy meer zou kunnen verhogen.

Discussion

The effects of self-assessment interventions on students' SRL

De resultaten suggereren dat zelfbeoordelingsinterventies een positief effect hebben op de strategieën van leerlingen voor zelfbeoordeling. Deze effecten variëren van klein tot middelgroot in de drie meta-analyses. De SRL meting leveren differentiële effecten op.

Differences between self-regulated measurements

Een belangrijke bijdrage van deze studie is het onderscheid dat wordt gemaakt tussen verschillende SRL-metingen. Hoewel de verschillende metingen van SRL in dezelfde richting wijzen, zijn er verschillen in de grootte van de effectgroottes.

Moderating effects on self-regulated learning

Geen van de variabelen had een statistisch significante invloed op de strategieën van de leerlingen. We interpreteren het resultaat voorzichtig en vermelden daarom alleen dat jongens en jongere leerlingen minder negatieve RLL-strategieën lijken te ervaren als gevolg van zelfevaluatie-interventies.

The effects of self-assessment interventions on students' self-efficacy



Zelfevaluatie-interventies hebben een aanzienlijke invloed op de zelfeffectiviteit van studenten. De bevindingen van de zelfeffectiviteitsanalyses impliceren dat zelfevaluatie-interventies een krachtige strategie kunnen zijn om het vertrouwen van studenten in hun eigen capaciteiten te vergroten. Dit is op zijn beurt weer belangrijk omdat zelfeffectiviteit een van de belangrijkste voorspellers is van de prestaties van studenten.

Moderating effects on self-efficacy

Twee van de vier geanalyseerde moderatoren hebben een statistisch significante invloed op de zelfeffectiviteit van studenten: geslacht en type zelfevaluatie-interventie. De bevindingen suggereren dat de zelfeffectiviteit van meisjes meer toeneemt als gevolg van een zelfevaluatie-interventie dan die van jongens.

Educational implications

Zelfevaluatie is een noodzaak voor productief leren. Zelfevaluatie blijkt een positieve invloed te hebben op academische prestaties en volgens de resultaten van ons onderzoek heeft zelfevaluatie ook een positieve invloed op de SRL-strategieën en de zelfeffectiviteit van studenten. Wij herhalen daarom dat zelfevaluatie een belangrijk instructie doel moet zijn in elke klas, maar niet alleen om de prestaties te verbeteren. Aangezien zelfevaluatie een impact heeft op de zelfredzaamheid en het zelfoplossend vermogen van leerlingen, moet zelfevaluatie worden geïmplementeerd in het belang van empowerment en zelfonderhoudend leren van leerlingen.

7.3 Carless & Boud (2018) The development of student feedback literacy: enabling uptake feedback

Feedbackvaardigheden van studenten duiden op het inzicht, de capaciteiten en die nodig zijn om informatie te begrijpen en te gebruiken om werk of leerstrategieën te verbeteren. d en een aantal belemmeringen voor het gebruik van feedback worden besproken. Vier onderling gerelateerde kenmerken worden voorgesteld als een kader dat de feedbackvaardigheden van studenten ondersteunt: feedback waarderen; oordelen; omgaan met affecten; en actie ondernemen.

Introduction

De krachtigste invloed op prestaties is feedback, maar de effecten zijn zeer variabel. Om het feedbackproces te verbeteren, moeten de leerlingen weten hoe feedback doeltreffend kan zijn en moeten zij de gelegenheid krijgen om feedback in het leerplan te gebruiken. Feedback wordt gedefinieerd als een proces waarbij leerlingen informatie uit verschillende bronnen op hun waarde weten te schatten en gebruiken om hun werk of leerstrategieën te verbeteren.

De focus van dit artikel ligt op de ontwikkeling van feedbackvaardigheden bij studenten in het hoger onderwijs als een manier om het opnemen van feedback door studenten mogelijk te maken. De implicaties van studentfeedback-geletterdheid voor onderwijs en cursusontwerp zijn niet voldoende overwogen. Een van de belangrijkste hindernissen voor effectieve feedback is over het algemeen lage niveaus van feedbackvaardigheid bij studenten.

Sociaal-constructivistische benaderingen richten zich op de onderlinge afhankelijkheid van sociale en individuele processen bij de co-constructie van kennis. Vier kenmerken van feedbackvaardigheden van studenten worden voorgesteld als een kader voor feedbackgeletterdheid bij studenten: feedback waarderen; oordelen; omgaan met affecten; en actie ondernemen. We bespreken hoe deze kenmerken kunnen worden ontwikkeld via twee activiteiten die in het curriculum zijn ingebed: peer feedback en discussie over voorbeelden.

Feedback processes and the student response

Studenten reageren op verschillende manieren op feedback binnen specifieke disciplines, curricula en contextuele settings; en in relatie tot hun eerdere ervaringen en hun eigen persoonlijke kenmerken.

Appreciating feedback processes

Het waarderen van feedback verwijst zowel naar studenten die de waarde van feedback erkennen als naar studenten die begrijpen hun actieve rol in het feedbackproces. Feedback-geletterdheid vereist dat leerlingen de academische taal verwerven die nodig is om te begrijpen, het interpreteren van en denken met complexe ideeën.

Making judgements

Om het meeste uit feedbackprocessen te halen, moeten studenten evaluatieve oordeelsvorming ontwikkelen: het vermogen om beslissingen te nemen over de kwaliteit van het werk van zichzelf en anderen (Tai et al. 2017). Lager presterende studenten zijn vaak relatief zwak in het zelf evalueren van hun prestaties en verwarren vaak inspanning met kwaliteit. Studenten hebben meer mogelijkheden voor zelfevaluatie nodig, zodat ze hun oordeel in de loop van de tijd kunnen verbeteren. Hun eigen oordeel vormden dat dit nodig is. Studenten betrekken bij het verbeteren van hun vermogen om goede oordelen te vellen is een uitdaging, tenzij er blijvende mogelijkheden zijn voor vergelijking met de opvattingen van anderen. Peer feedback is een belangrijk middel om het delen van oordelen door studenten aan te moedigen. Oordelen vellen houdt de impliciete of expliciete toepassing van criteria in.

Managing affect

Affect verwijst naar gevoelens, emoties en attitudes. Studenten reageren vaak defensief op feedback, vooral wanneer het commentaar kritisch is of de cijfers laag. Onder deze omstandigheden lukt feedback vaak negatieve affectieve reacties uit.

Een doel van feedback is om studenten uit te dagen om nieuwe perspectieven in te nemen, maar studenten zijn vaak niet maar studenten zijn vaak niet blij met deze uitdaging. Kritische feedback kan positieve of negatieve effecten hebben, afhankelijk van een reeks factoren, waaronder de zelfeffectiviteit, motivatie en het vermogen van studenten om constructief met emoties om te gaan.

Taking action

Feedbackvaardigheden vereisen dat leerlingen handelen naar aanleiding van de opmerkingen die ze hebben gekregen. Leerlingen moeten zich actief bezighouden met het begrijpen van informatie en deze gebruiken om hun latere werk te onderbouwen, en zo een feedback-lus sluiten. Om gebruik te kunnen maken van feedbackinformatie moeten studenten een repertoire aan strategieën hebben om productief te kunnen handelen.

Features of student feedback literacy

Studenten met een goed ontwikkelde feedbackvaardigheid waarderen hun eigen actieve rol in feedback processen; ontwikkelen voortdurend capaciteiten om een goed oordeel te vellen over academisch werk; en gaan op een positieve manier om met affecten. Deze drie kenmerken maximaliseren het potentieel voor studenten.

Enabling activities to develop student feedback literacy

Peer feedback en het analyseren van voorbeelden, worden besproken om te illustreren hoe ze explicieter kunnen worden gericht op de ontwikkeling van de feedbackvaardigheid van studenten.

Composing and receiving feedback

Bij peer feedback of peer review evalueren en oordelen leerlingen over het werk van medeleerlingen. Blootgesteld worden aan het werk van peers helpt leerlingen hun eigen

productie effectiever zelf te evalueren, omdat ze vergelijkingen maken tussen hun eigen werk en dat van anderen. Het geven van commentaar aan medeleerlingen is vaak gunstiger dan het ontvangen ervan omdat het cognitief-engagementiever is: het betrekken van hogere-orde processen, zoals het toepassen van criteria, het diagnosticeren van problemen en het aandragen van oplossingen peer review-processen vergemakkelijken de ontwikkeling van feedbackgeletterdheid bij studenten op verschillende manieren de verantwoordelijkheid voor feedback wordt bij de studenten gelegd; er zijn mogelijkheden voor dialoog; het proces gaat gepaard met het leren oordelen; en er is potentieel voor feedbacklussen die gesloten kunnen worden door actie van studenten. Zonder training en ondersteuning voor peer feedback is het onwaarschijnlijk dat de verwachte winst wordt behaald. peer review bevordert de feedbackvaardigheid van studenten wanneer studenten de waarde van peer feedback inzien en gecoacht worden in hoe het effectief uit te voeren.

Analysing exemplars

Leren van voorbeelden van verschillende genres maakt deel uit van de inductie in academische discoursen. Exemplaren zijn zorgvuldig gekozen voorbeelden van studentenwerk die worden gebruikt om dimensies van kwaliteit te illustreren en beoordelingsverwachtingen (Carless en Chan 2017). Door verwachtingen duidelijk te maken, verminderen voorbeeldwerken een deel van de onvermijdelijke ongerustheid van studenten over beoordelingsnormen en -vereisten. Doelgerichte analyse van voorbeelden ontwikkelt de feedbackgeletterdheid van studenten door hen te laten zien in plaats van hen te vertellen over kwaliteitswerk.

Kortom, door het analyseren van voorbeelden kunnen studenten beter inschatten hoe kwaliteit tot uiting komt, hun capaciteiten vergroten om een goed academisch oordeel te vellen, en de juiste conclusies trekken voor acties om hun eigen werk te verbeteren.

The teacher role

Activiteiten die feedback van leerlingen mogelijk maken, zullen waarschijnlijk alleen succes hebben bij de ontwikkeling van hun feedbackvaardigheden als leerkrachten geschikte leerplanomgevingen creëren voor de actieve deelname van leerlingen, en ook zorgen voor de nodige begeleiding, coaching en modellering. Een belangrijke rol van de leerkracht is de grondgedachte van activerende activiteiten over te brengen; uit te leggen hoe ze werken, de potentiële voordelen voor leerlingen uit te werken, en de uitdagingen aan te pakken waarmee leerlingen kunnen tegenkomen.

Het ontwerp van beoordelingen is een bijkomend kenmerk dat van invloed is op de vooruitzichten voor de ontwikkeling van feedbackvaardigheden van studenten. De manier waarop docenten beoordelingsstaken ontwerpen, opent of sluit mogelijkheden voor productieve feedbackprocessen.

Implications

Hieruit vloeien een aantal implicaties voor het onderwijs voort. Ten eerste is er behoefte aan meta-dialogen tussen docenten en studenten over feedback processen. Ten tweede moeten faciliterende activiteiten kernelementen van het curriculum worden om hun potentieel om de ontwikkeling van feedbackvaardigheden bij studenten te ondersteunen en evaluatieve oordeelsvorming te bevorderen. Ten derde zijn leeractiviteiten waarbij studenten samen feedback bespreken bijzonder nuttig,

Conclusion

Het argument dat in dit artikel wordt betoogd dat door de ontwikkeling van feedbackvaardigheden, leerlingen beter in staat zijn om informatie te gebruiken informatie te gebruiken om hun eigen werk te beoordelen en hun leren te verbeteren. Vier kenmerken van feedback vier kenmerken van feedbackvaardigheid bij studenten: feedbackprocessen waarderen; vaardigheden ontwikkelen om te oordelen; omgaan met affect; en actie ondernemen om feedback te gebruiken. De ontwikkeling van feedbackvaardigheden bij studenten staat



centraal bij het verbeteren van feedbackprocessen en bredere pogingen om de leerresultaten van studenten te verbeteren