

Diagnostisch rekenen-wiskunde assessment in groep 5

Diagnostisch assessment werd oorspronkelijk ingezet voor het diagnosticeren van speciale onderwijsbehoeften. Met de komst van Passend Onderwijs worden diagnostische instrumenten ontwikkeld om het formatief handelen van leraren te ondersteunen. In het hier beschreven promotieonderzoek werden de kenmerken van diagnostische instrumenten voor rekenen-wiskunde in groep 5 onderzocht.

De getallenlijn bleek geschikt om strategiegebruik van leerlingen te meten, maar was niet voor alle leerlingen geschikt. Daarnaast bleken sommige aftrekopgaven zoals $1000 - 680 = (3/4\text{-posities} - 3\text{ posities})$ het meest geschikt om overbruggingsfouten te diagnosticeren. Verrassend genoeg bleken vaardige groep-5-leerlingen meer overbruggingsfouten te maken dan laagvaardige leerlingen. Dit impliceert dat diagnostische instrumenten ook geschikt kunnen zijn om het formatief handelen voor leerlingen zonder reken-wiskunde problemen te ondersteunen. Kortom, diagnostische instrumenten hebben de potentie misconcepties vroegtijdig te signaleren.

Inleiding

In het onderwijs verwijst assessment naar het proces van verzamelen van kwalitatieve en/of kwantitatieve data over het leerproces en/of de leerprestaties van leerlingen (Stobart, 2008). Diagnostisch assessment werd oorspronkelijk ingezet voor het diagnosticeren van speciale onderwijsbehoeften met als doel te bepalen of leerlingen verwezen moesten worden naar het speciaal onderwijs (Bejar, 1984; Vermeulen, 2021). Door beleid zoals 'Weer samen naar school' en 'Passend onderwijs' (Ledoux et al., 2020) bleven meer leerlingen met speciale onderwijsbehoeften in het reguliere onderwijs. Hierdoor verschoof de doelstelling van diagnostisch assessment naar afstem-

Jorine Vermeulen, Cito

Vermeulen, B. (2024). Diagnostisch rekenen-wiskunde assessment in groep 5. *Volgens Bartjens – ontwikkeling en onderzoek*, 43(4), 56-61.

ming op de onderwijsbehoeften van alle leerlingen in het reguliere onderwijs. Deze cyclische afstemming wordt ook formatief handelen genoemd.

Diagnostisch assessment is een cyclisch proces gericht op het vaststellen van domein-specifieke onderwijsbehoeften. Met diagnostische instrumenten wordt binnen een klein onderdeel van de leerlijn ingezoomd op de conceptuele en procedurele ontwikkeling van individuele leerlingen (Leighton & Gierl, 2007; Vermeulen, 2021). In het hier beschreven promotieonderzoek zoomde ik in op de leerlijn van rekenen-wiskunde in groep 5. Diagnostische taken worden gebruikt om het reken-wiskundig denken van leerlingen uit te lokken. Vervolgens wordt een diagnostisch raamwerk gebruikt om het antwoordgedrag van leerlingen op diagnostische taken te analyseren en te interpreteren in termen van onderwijsbehoeften. Een diagnostisch raamwerk is gebaseerd op leerpsychologische en didactische inzichten over rekenstrategieën en veel voorkomende misconcepties en fouten. Zo ben ik op basis van internationale visies op reken-wiskunde ontwikkeling (zie bijvoorbeeld Rittle-Johnson, 2017), uitgegaan van een iteratieve relatie tussen procedurele en conceptuele kennis en vaardigheden.

Ik onderzocht de kenmerken van diagnostische instrumenten die leraren in groep 5 ondersteunen bij het formatief handelen binnen het aftrekken tot 1000. Hierbij stonden de volgende onderzoeksvragen centraal:

1. Welk type antwoordgedrag is diagnostisch relevant voor formatieve beslissingen bij rekenen-wiskunde in groep 5?
2. Welke kenmerken moeten diagnostische taken hebben om antwoordgedrag dat relevant is voor formatieve beslissingen te verzamelen?

Het type antwoordgedrag heeft in dit onderzoek betrekking op het oplossingsproces (dat wil zeggen: strategiegebruik) en het (foute) antwoord van de leerling. Om de onderzoeksvragen te beantwoorden voerde ik een pilotstudie en twee experimenten uit. De resultaten van de verschillende experimenten zijn in vijf deelstudies beschreven. De pilotstudie en het eerste experiment richtte zich op het gebruik van papieren en digitale getallenlijnen voor het diagnosticeren van strategiegebruik (hoofdstuk 2 t/m 4 in Vermeulen (2021)). Het tweede experiment richtte zich op het signaleren van systematische fouten en gerelateerde misconcepties binnen het aftrekken tot 1000. Ik ontwierp en onderzocht aftrekopgaven die over het tien- en honderdtal gaan en evalueerde de mate waarin zij de beoogde fouten konden uitlokken (hoofdstuk 5 en 6 in Vermeulen (2021)). Omdat ik de resultaten van de deelstudies beschrijf geef ik een korte samenvatting van de theoretische achtergrond van het promotieonderzoek.

Noodzaak voor formatieve diagnostische instrumenten

Internationaal worden er verschillende begrippen gebruikt voor het gebruik van assessment data voor de afstemming tussen het leerstofaanbod en de waargenomen ontwikkeling van leerlingen. De meest gebruikte termen in Nederland zijn 'Handelingsgericht werken' (HGW) en 'Opbrengstgericht werken' (OGW). Deze benaderingen van formatief assessment verschillen in de typen data die worden verzameld en de soorten beslissingen die over de inrichting van het onderwijs worden genomen (Van der Kleij et al., 2015). HGW is meer gericht op de behoeften van de klas en individuele leerling, terwijl OGW zich meer richt op de evaluatie van de onderwijskwaliteit op (boven)schools- en klasniveau. Assessmentinstrumenten zoals leerlingvolgsysteem (LVS-)toetsen kunnen zowel voor HGW als voor OGW worden gebruikt. Toetsresultaten van LVS-toetsen en methodetoetsen geven echter weinig specifieke informatie over de onderwijsbehoeften van leerlingen. Dergelijke toetsen geven vooral informatie over welke doelen zijn bereikt en welke nog niet. Om te bepalen welke instructie en/of oefenstof leerlingen nodig hebben om eerder aangeboden en/of nieuwe doelen te bereiken heeft een leraar specifiekere informatie nodig. Zoals eerder genoemd worden diagnostische instrumenten ontwikkeld vanuit een diagnostisch raamwerk waarin beschreven staat hoe een specifieke vaardigheid – zoals aftrekken tot 1000 – zich ontwikkelt. Veel voorkomende fouten en misconcepties maken deel uit van zo'n diagnostisch raamwerk. Ook de beschrijvingen van het reken-wiskundig denken en handelen volgens het handelingsmodel kunnen deel uitmaken van zo'n diagnostisch raamwerk.

Rekenspecialisten, ib'ers en leraren maken vaak bewust of onbewust gebruik van diagnostische raamwerken als ze een (diagnostisch) rekengesprek met leerlingen voeren om de onderwijsbehoeften van een leerling vast te stellen. Dergelijke gesprekken vinden echter voornamelijk plaats als er sprake is van (tijdelijke) reken-wiskunde problemen en worden niet met alle leerlingen gevoerd (Veldhuis et al., 2013). Diagnostische instrumenten maken gebruik van taken die – in principe –

voor alle leerlingen in te zetten zijn. In dat opzicht zijn ze vergelijkbaar met methode- en LVS-toetsen. Omdat de taken in het diagnostische instrument zijn gebaseerd op een diagnostisch raamwerk is het mogelijk om strategiegebruik en/of misconcepties bij groepen leerlingen vast te stellen. Kortom, diagnostische instrumenten zijn een noodzakelijk onderdeel van formatief assessment. In de volgende paragraaf geef ik een korte samenvatting van de verschillende deelstudies waarin de lege getallenlijn het belangrijkste kenmerk van het diagnostische instrument was.

Diagnostisch raamwerk 1: strategiegebruik en lege getallenlijn

Tot groep 4 ligt de nadruk op het lineaire getalbegrip: vanuit het leren tellen leren leerlingen dat getallen geordend kunnen worden (op een getallenlijn) van klein naar groot. Leren optellen en aftrekken krijgt onder andere de betekenis van sprongen maken op die (denkbeeldige) getallenlijn (Van den Heuvel-Panhuizen, 2008; Teppo & Van den Heuvel-Panhuizen, 2013). Dit lineaire getalbegrip gaat samen met het leren rijgen, bijvoorbeeld, $83 - 57 =$ via $83 - 20 = 63$; $63 - 20 = 43$; $43 - 10 = 33$; $33 - 3 = 30$; $30 - 4 = 26$. Voor het rijgen tot 100 moeten leerlingen begrijpen dat getallen gesplitst kunnen worden. Kraemer (2011) deed onderzoek naar strategiegebruik in groep 5 en ontwierp een classificatiesysteem waarin hij onderscheid maakte tussen verschillende niveaus van rijgen. Vergelijkbaar met het handelingsmodel (Van Groenestijn, et al., 2011) beschreef Kraemer (2011) dat leerlingen steeds kortere, abstractere en formele manier van rijgen toepassen. Kraemer (2011) ontwikkelde bij Cito een diagnostisch instrument waarmee leraren het strategiegebruik konden uitlokken en classificeren. Dit diagnostische instrument bestond uit taken en een hard-op-denkprotocol waarmee leraren leerlingen konden interviewen en observeren. Hard-op-denkprotocollen hebben als nadeel dat het voor leerlingen moeilijk is om hun strategiegebruik onder woorden te brengen. Bovendien hebben leraren geen tijd om alle leerlingen individueel te onderzoeken. Daarom onderzocht ik de mate waarin de lege getallenlijn voor groep-5 leerlingen geschikt is als onderdeel van een diagnostisch instrument dat klassikaal afgenomen zou kunnen worden.

De pilotstudie was een eerste - hoofdzakelijke kwalitatieve en informele - verkenning van de diagnostische mogelijkheden van de getallenlijn. Bij twee groepen 5 van twee verschillende basisscholen verzamelde ik voor tien optel- en aftrekepgaven tot de 100 van 30 leerlingen 600 getallenlijnen. Aan de hand van de literatuur en de kwalitatieve analyse van de 600 getallenlijnen ontwierp ik een diagnostisch raamwerk voor het diagnosticeren van procedurele en conceptuele kennis van rijgen tot 100. Tijdens de pilot bleek dat niet alle leerlingen in groep 5 het fijn vonden om *verplicht* te rekenen op de lege getallenlijn. Ook bleek dat sommige opgaven te makkelijk waren en de getallenlijn eigenlijk overbodig was. Bovendien was het erg veel werk om 600 getallenlijnen te analyseren. Tot slot waren de twee leraren van de klassen die deelnamen aan de pilot niet overtuigd van de diagnostische capaciteit van de lege getallenlijn.

Op basis van de ervaringen in de pilot ontwierp ik in het eerste experiment een tabletapplicatie G3T4ll#NL1JNapp. Tablettechnologie heeft als voordeel dat het handschrift van leerlingen niet ontcijferd hoeft te worden en de stappen op de getallenlijn chronologisch worden opgeslagen. Het gebruik van tablets kan echter ook het antwoordgedrag van leerlingen beïnvloeden, zeker omdat leerlingen tot dan toe vooral gewerkt hebben met getallenlijnen op papier.

Op twee momenten zijn 123 groep-5-leerlingen van negen basisscholen getoetst met een papieren of tablet optel- en aftrekepgaven waarbij de getallenlijn vrijwillig gebruikt kon worden. Om het vrijwillig gebruik van de getallenlijn te stimuleren zijn slechts aftrek- en optelopgaven in het getalsdomein tot 1000 opgenomen die het tiental overschrijden. Uit de pilot was immers gebleken dat als opgaven te gemakkelijk zijn de getallenlijn overbodig is.

Met covariantieanalyses is zowel naar verschillen tussen als binnen de populatie leerlingen gekeken. Hieruit bleek dat leerlingen in de tabletconditie vaker verkorte hoofdrekstrategieën gebruikten dan de leerlingen in de papieren conditie. Wat betreft de verschillen binnen de populatie leerlingen bleek dat leerlingen die tijdens het eerste afnamemoment de papieren taak maakten, tijdens het tweede afnamemoment minder vaak de getallenlijn gebruikten in de tabletaak. Aanvullend bleek dat leerlingen die tijdens het eerste afnamemoment de papieren taak maakten tijdens het tweede afnamemoment op de tablet significant meer fouten maakten. Kortom, de tablettechnologie bleek het strategiegebruik van leerlingen negatief te beïnvloeden. Ook de frequentie waarmee leerlingen de getallenlijn gebruikten viel tegen. Op basis van de pilot en dit experiment met de G3T4ll3NL1JNapp moesten we concluderen dat vrijwillige gebruik van de lege getallenlijn en tablets niet voor alle leerlingen in groep 5 geschikt zijn. Daarom ontwierpen we een tweede diagnostisch kader gericht op splitsend aftrekken in het getalsdomein tot 1000.

Diagnostisch raamwerk 2: misconcepties splitsend aftrekken tot 1000

Door de uitbreiding van het getalsdomein naar 1000 komt er in groep 5 meer nadruk te liggen op het decimaal-positionele getalbegrip. Getallen bestaan uit cijfers en de positie van het cijfer in het getal bepaalt de waarde: de '3' in 354 is 300 waard (Kraemer, 2011; Fuson, 1990; Fuson et al., 1997). Parallel aan het uitbreiden van het getalbegrip leren leerlingen splitsend rekenen. Splitsend rekenen houdt in dat zowel de 57 als de 83 worden gesplitst in tientallen en eenheden: $80 - 50 = 30$; $3 - 7 = 4$ tekort, $30 - 4 = 26$. Dit voorbeeld laat een potentiële misconceptie van leerlingen in groep 5 zien: leerlingen die nog niet volledig hoe het decimaal-positioneel rekenen werkt draaien $3 - 7$ om naar $7 - 3 = 4$ waardoor zij $83 - 57 = 34$ antwoorden. Het gaat hier om een overbruggingsfout. Deze overbruggingsfout staat in de internationale literatuur bekend als de 'smaller-from-larger error' (Brown & Burton, 1978; Resnick, 1982). In de praktijk wordt deze fout ook wel de omwisselfout genoemd. In het tweede experiment ging ik uit van drie soorten overbruggingsfouten:

Bijvoorbeeld $43 - 17 =$

Omwisselfout $40 - 10 = 30$, $7 - 3$ (in plaats van $3 - 7$) = 4 , $30 + 4 = 34$.

Omwisselfout met tiental of honderdtal compensatie: $40 - 10 = 30$, merkt op dat $3 - 7$ het inwisselen van een tiental vergt: $30 - 10 = 20$, toepassen $7 - 3$ (in plaats van $3 - 7$) = 4 , $20 + 4 = 24$.

Geen omwisselfout, maar vergeten het tiental of honderdtal te compenseren. $40 - 10 = 30$, $(1)3 - 7 = 6$, $30 + 6 = 36$.

Studie 1: diagnostische capaciteit meercijferige aftrekopgaven

In de eerste studie bij dit tweede experiment heb ik de diagnostische capaciteit van meercijferige aftrekopgaven geëvalueerd. Diagnostische capaciteit werd gedefinieerd als de mate waarin meercijferige aftrekopgaven geschikt zijn om overbruggingsfouten te ontlokken. Het ging hierbij om opgaven zoals '1000 - 680 =' waarvoor overbrugging van het tiental of honderdtal noodzakelijk was. De itemkenmerken die zijn onderzocht waren: open en meerkeuze items, kale en context items (zonder afbeelding) en verschillende getalskenmerken (afbeelding 1). Itemresponstheorie is gebruikt om de itemparameters te schatten. Vervolgens zijn deze itemparameters gebruikt in twee variantieanalyses om de diagnostische capaciteit van open versus meerkeuzeopgaven, kale versus contextopgaven en drie categorieën van getalskenmerken te vergelijken. Zoals verwacht bleken meerkeuze items een hogere diagnostische capaciteit te hebben dan open items. Interessanter was echter het resultaat dat het aantal positiewaarden (n) in het aftrektaal en de aftrekker de diagnostische capaciteit bepaalde. Items uit de categorie $3/4n - 3n$, zoals $1000 - 680 =$ hadden de hoogste diagnostische capaciteit. Daarentegen hadden items uit de categorie $3/4n - 2n$, zoals $1000 - 20 =$ de laagste diagnostische capaciteit.

► Itemtypes voor overbruggingsfouten in aftrekken tot 1000

Type	Lenen van	Positiewaarden	Voorbeelditem
1a	10	$2nb - 2n$	$83 - 26 =$
2	10	$3n - 3n$	$453 - 127 =$
3	100	$3n - 2n$	$347 - 62 =$
4	100	$3n - 3n$	$634 - 251 =$
5	10	$2n - 2n =$	$70 - 43 =$
6	100	$3n - 2n =$	$406 - 22 =$
7a	10, 100	$3n - 2n =$	$400 - 27 =$
8	100, 1000	$1000 - 2n =$	$1000 - 70 =$
9a	100, 1000	$1000 - 3n =$	$1000 - 340 =$

^a Types die zijn gekloond van de tweede generatie LVS-toetsen in groep 5.

^b Aantal posities binnen het getal.

Studie 2: relatie overbruggingsfouten en rekenvaardigheid

In de tweede studie naar overbruggingsfouten onderzocht ik de relatie tussen overbruggingsfouten en rekenvaardigheid. In de studie participeerden 694 groep-5-leerlingen en 35 leraren van 25 Nederlandse basisscholen. Multilevel regressieanalyses lieten zien dat - na controle voor het totaal aantal fouten - het aantal overbruggingsfouten positief samenhangt met de rekenvaardigheid. Dit betekent dat leerlingen die verhoudingsgewijs meer overbruggingsfouten dan andere fouten maak-

ten gemiddeld genomen een hogere rekenvaardigheid hadden. Verrassend genoeg impliceert dit resultaat dat het diagnosticeren van overbruggingsfouten vooral geschikt is voor leerlingen met een bovengemiddelde rekenvaardigheid. De diagnose van overbruggingsfouten geeft informatie over waar in hun procedurele en conceptuele ontwikkeling binnen aftrekken deze leerlingen staan.

Conclusies

Het hier beschreven onderzoek richtte zich voornamelijk op de leerlijn aftrekken tot 1000, de lege getallenlijn, tablettechnologie en overbruggingsfouten. Met de hiervoor beschreven deelstudies heb ik een bijdrage geleverd aan het antwoord op de twee onderzoeksvragen die in mijn proefschrift centraal stonden.

1. Welk type antwoordgedrag is diagnostisch relevant voor formatieve beslissingen in reken-wiskunde in groep 5?
2. Welke kenmerken moeten diagnostische taken hebben om antwoordgedrag dat relevant is voor formatieve beslissingen te verzamelen?

Het meest verrassende resultaat van het hier beschreven promotieonderzoek is dat sommige misconcepties – zoals overbruggingsfouten – vaker voorkomen bij rekenvaardige leerlingen. Dat betekent dat het zinvol is om niet alleen diagnostische instrumenten te ontwikkelen voor leerlingen met milde en/of ernstige reken-wiskunde problemen. Op basis van dit onderzoek kan voor de eerste vraag geconcludeerd worden dat vanuit een theoretische didactisch oogpunt zowel strategiegebruik als fouten relevant zijn. Hierbij gaat het met name om de relatie tussen conceptuele en procedurele kennis. Bij zowel de getallenlijnoplossingen als bij de overbruggingsfouten ging het om de interactie tussen conceptuele en procedurele kennis die leidde tot bepaald antwoordgedrag op de diagnostische taken. Voor de tweede vraag kan dus gezegd worden dat diagnostische taken in staat moeten zijn om antwoordgedrag uit te lokken waaruit afgeleid kan worden welke procedurele en/of conceptuele kennis leidde tot dit antwoordgedrag. Bij het bepalen van de kenmerken van diagnostische taken moet niet alleen oog zijn voor de reken-wiskundige kenmerken zoals de getallen, maar ook naar technologische kenmerken. Ondanks dat technologie kansen biedt voor het geautomatiseerd scoren van antwoordgedrag, blijkt uit mijn onderzoek dat technologische kenmerken het antwoordgedrag negatief kunnen beïnvloeden. Deze negatieve invloed kan ten koste gaan van de validiteit van het diagnostische instrument.

Noten

¹ Dit instrument wordt niet langer uitgegeven.

² De applicatie is geprogrammeerd door Patrick de Klein (werkzaam bij CitoLab).

Literatuur

- Bejar, I. I. (1984). Educational diagnostic assessment. *Journal of Educational Measurement*, 21(2), 175–189. <https://doi.org/10.1111/j.1745-3984.1984.tb00228.x>
- Brown, J., & Burton, R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science*, 2(2), 155–192. [https://doi.org/10.1016/S0364-0213\(78\)80004-4](https://doi.org/10.1016/S0364-0213(78)80004-4)
- Kraemer, J. M. (2011). *Oplossingsmethoden voor aftrekken tot 100 [solution methods for subtraction up to 100]* [Doctoral dissertation, Technical University Eindhoven]. <https://doi.org/10.6100/IR721544>
- Fuson, K. C. (1990). Conceptual Structures for Multiunit Numbers: Implications for Learning and Teaching Multi-digit Addition, Subtraction, and Place Value. *Cognition and Instruction*, 7(4), 343–403. https://doi.org/10.1207/s1532690xci0704_4
- Fuson, K. C., Wearne, D., Hiebert, J. C., Murray, H. G., Human, P. G., Olivier, A. I., Carpenter, T. P., & Fennema, E. (1997). Children's conceptual structures for multi-digit numbers and methods of multi-digit addition and subtraction. *Journal for Research in Mathematics Education*, 28(2), 130–162. <http://www.jstor.org/stable/10.2307/749759>
- van Groenestijn, M., Borghouts, C., Janssen, C., van Groenestijn, M., Borghouts, C., & Janssen, C. (2011). *Protocol Ernstige RekenWiskunde-Problemen en Dyscalculie*. Van Gorcum. https://erwd.nl/_downloads/protocol-ernstige-reken-wiskunde-problemen-en-dyscalculie/basisonderwijs/protocol-erwd-po-bso-so.pdf
- van den Heuvel-Panhuizen, M. (2008). Learning from “Didactikids”: An impetus for revisiting the empty number line. *Mathematics Education Research Journal*, 20(3), 6–31. <http://www.springerlink.com/index/R3212763N95PLOV0.pdf>
- van der Kleij, F.M., Vermeulen, J.A., Schildkamp, K., Eggen, T.J.H.M. (2015). Integrating data-based decision making, assessment for learning, and diagnostic testing in formative assessment. *Assessment in Education: Principles, Policy & Practice*, 22, 324–343, doi:10.1080/0969594X.2014.999024
- Ledoux, G., & Waslander, S., m.m.v. Eimers, T. (2020). Evaluatie Passend Onderwijs. Eindrapport. Kohnstamm Instituut. <https://evaluatiepassendonderwijs.nl/wp-content/uploads/2020/05/Eindrapport-evaluatie-passend-onderwijs-2020.pdf>
- Leighton, J. P., & Gierl, M. J. (Eds.) (2007). *Cognitive Diagnostic Assessment for Education: Theory and Applications*. New York, NY: Cambridge University Press.

- Resnick, L. B. (1982). *Syntax and Semantics in learning to subtract*. [report] Pittsburgh Univ., PA. Learning Research and Development Center. <https://eric.ed.gov/?id=ED221386>
- Rittle-Johnson, B. (2017). Developing Mathematics Knowledge. *Child Development Perspectives*, 11(3), 184–190. <https://doi.org/10.1111/cdep.12229>
- Stobart, G. (2008). *Testing times: The uses and abuses of assessment*. Routledge.
- Teppo, A., & van den Heuvel-Panhuizen, M. (2013). Visual representations as objects of analysis: the number line as an example. *ZDM*, 46(1), 45–58. <https://doi.org/10.1007/s11858-013-0518-2>
- Veldhuis, M., van den Heuvel-Panhuizen, M., Vermeulen, J. A., & Eggen, T. J. H. M. (2013). Teachers' Use of Classroom Assessment in Primary School Mathematics Education in the Netherlands. *CAD-MO*, 35–53. doi:10.3280/CAD2013-002004
- Vermeulen, J. A. (2021). *Diagnostic Mathematics Assessment in the Third grade*. [Dissertation]. University of Twente. DOI: 10.3990/1.9789036552295

Initially, diagnostic assessment was used for diagnosing special educational needs en referrals to special education. Due to inclusive educational policies, diagnostic instruments are being designed to facilitate teachers' formative practices. In this dissertation various characteristics of diagnostic instruments for third grade mathematics were evaluated. The empty number line was suitable for eliciting strategy use for some – but not all – students. Additionally, subtraction problems like 1000 – 680 (3/4-digits – 3 digits) were found to have the highest diagnostic capacity for diagnosing bridging errors. Surprisingly, mathematically highly skilled third grade students proportionally made more bridging errors compared to less skilled students. This implies that diagnostic instruments can also be used to advance the learning of students without apparent mathematical problems.