

## Appendix B: Tools voor kritische beoordeling van de methodologische kwaliteit

### Kritische beoordeling reviews

<b>Titel</b>	
<b>Auteur(s)</b>	
<b>Jaartal</b>	

<b>AANTAL 'JA' = #/14 (SR) OF #/15 (MA)</b>	2x niet duidelijk = 1 * NEEN
<b>AANTAL 'NEEN' =</b>	
<b>Let op [Afspraken]: Voor bovenstaande telling neem je:</b> <ul style="list-style-type: none"> <li>• bij '2b' de 'disagreement methode' NIET mee</li> <li>• bij '2d' slechts 1* JA of NEEN mee (zie toelichting bij '2d')</li> <li>• het eindkader (laatste pp. onder rode streep niet meenemen in telling)</li> </ul>	

<b>BESLISSING</b>	
• 0 - 2 * NEEN = HK	<input type="checkbox"/> HK= HOGE KWALITEIT
• 3 - 5 * NEEN = MK	<input type="checkbox"/> MK= GEMIDDELDE KWALITEIT
• >5 * NEEN = UIT	<input type="checkbox"/> UIT= LAGE KWALITEIT

<b>1/ ASSESSING RELEVANCE</b>		
Is de onderzoeksvraag duidelijk gesteld? Beschrijven de auteurs de in-/exclusiecriteria op een gepaste manier?	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
<b>2/ STUDY ELIGIBILITY</b>		
<b>2.a.</b> Is de zoekprocedure om relevante studies te identificeren duidelijk omschreven? ( <i>protocol of uitgeschreven – 'to support replication'</i> )	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
<b>2.b.</b> Werd de beslissing om studies te includeren gebaseerd op 2 of meer reviewers (onafhankelijk of gedeeltelijk) en werd er een methode beschreven in geval van disagreement?	<b>Beslissing:</b> <input type="checkbox"/> JA <input type="checkbox"/> NEEN <input type="checkbox"/> NIET DUIDELIJK	<b>Disagreement methode beschreven:</b> <input type="checkbox"/> JA <input type="checkbox"/> NEEN

<b>2.c. Is er een vorm van kwaliteitsbeoordeling van de opgenomen studies?</b>			
a. Werd er duidelijk omschreven welke criteria gehanteerd werden om de kwaliteitsbeoordeling van de studies aan te pakken (~replication)?	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN	
b. Werden enkel studies geïnccludeerd die statistisch analyseerbaar waren (o.m., pretest-gegevens zodat bv. ten minste ES berekend kon worden)?	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN	
c. Werd de Risk of Bias van/binnen de individuele studies nagegaan/beschreven? <i>(meest relevante)</i> Was er een check van ...			
~ baseline equivalentie (E vs.C): is de baseline equivalentie (gemiddelde pre-test verschillen E/C) meegenomen, ook bij evt. verlies van participanten?	<input type="checkbox"/> JA -RCT <input type="checkbox"/> JA - binnen marge 50% van SD <input type="checkbox"/> JA -opgenomen in moderator analyse	<input type="checkbox"/> NEEN	<input type="checkbox"/> NIET DUIDELIJK  <i>bv.</i> -the study <u>accounted for</u> pre-test differences -The investigators of the study <u>had to establish equivalence</u> of the experimental and comparison groups for quasiexperimental studies.
~ PERFORMANCE bias: <i>Werd er gebruik gemaakt van gestandaardiseerde of niet-gestandaardiseerde tests met wel een zekere kwaliteitsindicatie?</i>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN	<input type="checkbox"/> NIET DUIDELIJK
~ ATTRITION bias: <ul style="list-style-type: none"> <li>Ging de auteur van de MA/SR de attrition rate in de primaire studies na?</li> <li>Indien er een hoge attrition was, werd dan (1) OF de studie uitgesloten (2) OF de baseline equivalence op de overgebleven sample aangetoond (3) OF werd</li> </ul>	<input type="checkbox"/> JA  <input type="checkbox"/> JA (1)-(2)-(3)	<input type="checkbox"/> NEEN  <input type="checkbox"/> NEEN	<input type="checkbox"/> NIET DUIDELIJK

de attrition werd meegenomen als moderator.			
<b>2.d.</b> Werden de belangrijkste karakteristieken van de geïnccludeerde studies beschreven in een tabel of aan de hand van 'ranges'?	<input type="checkbox"/> JA (indien > 4 van 'a-h') <input type="checkbox"/> NEEN (indien ≤ 4 van 'a-h')		
a. Type interventie	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
b. Leeftijd of klas of onderwijsniveau	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
c. Duurtijd interventie	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
d. Uitvoerder interventie	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
e. Research designs	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
f. Sample size(s)	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
g. Uitkomstmaten (bv. woordenschat/fluency/...)	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN

h. Resultaten	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
Optioneel:			
i. Kenmerken deelnemers (meertaligheid/risicostatus/...)	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
j. Settings (kleine groep/grote groep/...)	<input type="checkbox"/> JA – ranges	<input type="checkbox"/> JA – tabel	<input type="checkbox"/> NEEN
<b>2.e.</b> Indien er melding wordt gemaakt van excluded studies, wordt er dan een reden gegeven waarom deze studies alsnog geëxcludeerd werden? <i>(studies die binnen protocol passen maar toch op het einde nog UIT gaan).</i>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN	<input type="checkbox"/> NIET VAN TOEPASSING
<b>3/ DATA COLLECTION</b>			
<b>3.a.</b> Gebeurde de ‘data extraction’ geheel of gedeeltelijk op basis van ten minste 2 reviewers?	<input type="checkbox"/> JA		<input type="checkbox"/> NEEN (ook indien het niet duidelijk is)
<b>3.b.</b> Werden de effect sizes van de individuele studies berekend en gerapporteerd (via Hedges’ g of Cohen’s d)? <i>(Cohen’s d en Hedges’g ‘berekenen’ eigenlijk hetzelfde, maar Hedges’g is accurater wanneer het gaat over kleine sample size (deze laatste corrigeert dus voor sample size))</i>	<input type="checkbox"/> JA, via Cohen’s d <input type="checkbox"/> JA, via Cohen’s d en met aanvulling Hedges’ g <input type="checkbox"/> JA, via Hedges’ g <input type="checkbox"/> JA, maar op andere manier: specifieer:		<input type="checkbox"/> NEEN
<b>3.c.</b> Werd de risk of ‘publication’ bias nagegaan?	<input type="checkbox"/> JA , maar hoe niet duidelijk <input type="checkbox"/> JA, en duidelijk welke methode (bv. funnel plots, the Begg rank correlation test, Egger regression, Duval and Tweedie trim-		<input type="checkbox"/> NEEN

	and-fill test, and Rosenthal failsafe-N test ...).	
<b>3.d.</b> Indien er een 'publication bias' was, werd deze dan OF aangepakt (bv. via trim & fill methode) OF ten minste meegenomen in de discussie bv. limitations?	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
		<input type="checkbox"/> NIET VAN TOEPASSING
<b>4/ Voor MA (statistisch combineren van de studies)</b>		
<b>4.a.</b> Indien er een overkoepelende ES werd berekend, werden dan mogelijk beïnvloedende factoren van de individuele studies (bv. sample sizes) meegenomen <i>~in lijn met wat Slavin ook aangeeft: analyses of combined effect sizes should control for sample size</i>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
		<input type="checkbox"/> NIET DUIDELIJK bv. Comprehensive Meta-Analysis software (Borenstein, et al., 2005) was used to carry out all statistical analyses such as Q statistics and overall effect sizes.
<b>NOTEREN MAAR NIET MEE NEMEN IN TELLING #NEEN</b>		
<b>AD1</b> Werd de heterogeniteit (heterogeneity) (of vanuit de andere kant bekeken homogeniteit (homogeneity)) van de ES's nagegaan (via Q-statistics of I <sup>2</sup> )? <i>A value greater than 25% is considered to reflect low heterogeneity, 50% moderate, and 75% high heterogeneity. However, there is no definitive cut-off value at which no data pooling should be performed.</i>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
<b>AD2</b> Welk soort model/analyse werd toegepast?	<input type="checkbox"/> fixed eff. model <input type="checkbox"/> random eff. model <input type="checkbox"/> andere-noteer:	<input type="checkbox"/> NEEN

<b>DUS</b> Combineren de auteurs aangetoonde heterogeniteit met een fixed effect model?	<input type="checkbox"/> JA- red flag!!!	
---	--	--

# Critical appraisal Effectstudie met cluster-level assignment

Dit sjabloon kies je wanneer de experimentele en controlegroep werden gevormd op basis van clusters. Bijvoorbeeld, klasgroepen werden toegewezen aan de interventie of controleconditie. Het kan hier zowel gaan over quasi-experimentele studies als RCT's.

<b>Titel</b>	
<b>Auteur(s)</b>	
<b>Jaartal</b>	

<p><b>SCORE</b></p> <p><b>DEEL 1:</b> 0 / -2</p> <p><b>DEEL 2:</b> /7</p> <p><b>Totaalscore</b> /7</p>	<p><b>DEEL 1</b></p> <p><b>1</b> = Cluster-RCT's die aan volgende 3 voorwaarden voldoen: lage attritie op niveau cluster + geen risk of bias door joiners + geen risk of bias door non-response of individuals</p> <p><b>-2</b> = Alle andere designs</p> <p><b>DEEL 2</b></p> <p>Elke JA is een punt. 2 x GEEN INFO is een punt.</p>
--	---

<b>BESLISSING</b>	
<p>- <math>\geq 4</math> = HK</p> <p>- <math>&lt; 4</math> = MK</p> <p>- STOP in DEEL 1 = LK</p>	<p><input type="checkbox"/> HK= HOGE KWALITEIT</p> <p><input type="checkbox"/> MK= MIDDELMATIGE KWALITEIT</p> <p><input type="checkbox"/> UIT= LAGE KWALITEIT</p>

<b>DEEL 1 Vereiste kwaliteitsindicatoren</b>		
<b>0/ CONFOUNDING FACTORS</b>		
<p><b>0A</b> Does the intervention or comparison group contain a <b>single unit</b> (n=1)?</p> <p><i>The intervention or comparison group contains a single study unit—such as a teacher, classroom, or school—and that unit is not present in the other condition.</i></p> <p><i>Bijvoorbeeld: two schools are randomly assigned, one to each condition; a study has two intervention classrooms and two comparison classrooms, but both intervention classrooms had the same teacher, who had no interaction with the comparison classrooms.</i></p>	<p><input type="checkbox"/> JA (STOP)</p>	<p><input type="checkbox"/> NEEN (ga naar 0B)</p>
<p><b>0B</b> Do the characteristics of the subjects in the intervention or comparison group differ systematically, <b>with no overlap</b>, in ways that are associated with the outcomes?</p> <p><i>Bijvoorbeeld: Teacher education level: a small group of teachers in a master's program implements the intervention, whereas students in the comparison group are taught by teachers with bachelor's degrees.</i></p> <p><i>Different time period (bijv. successive-cohort designs)</i></p> <p><i>Student level: Intervention students are all in grade 8, and comparison students are all in grade 7.</i></p>	<p><input type="checkbox"/> JA (STOP)</p>	<p><input type="checkbox"/> NEEN (ga naar 1A)</p>

<i>Student characteristics: Intervention students are all English learners, and no comparison students are.</i>		
<b>1/ LEVEL OF ANALYSIS</b>		
<b>1A</b> Gebeurde de <b>analyse</b> van de resultaten op het <b>niveau</b> van de cluster?	<input type="checkbox"/> JA (ga naar 2)	<input type="checkbox"/> NEEN (ga naar 1B)
<b>1B</b> Indien de auteurs de resultaten op het niveau van het individu analyseerden, gebruikten ze in hun statistische berekeningen dan methodes die voor deze ‘ <b>unit of analysis error</b> ’ corrigeren?	<input type="checkbox"/> JA (ga naar 2)	<input type="checkbox"/> NEEN (STOP)
<b>2/ RANDOM ASSIGNMENT &amp; CLUSTER-LEVEL ATTRITION</b>		
<b>2</b> Is de studie een cluster-RCT met lage attritie op <b>clusterniveau</b> ? <i>Cluster-level attrition: het verlies van (een) volledige cluster(s)</i> <i>Berekening = zelfde werkwijze als attrition op niveau van individu</i>	<input type="checkbox"/> JA (ga naar 3)	<input type="checkbox"/> NEEN (ga naar 5) <input type="checkbox"/> GEEN INFO (ga naar 5)
Berekening		
	N =	Attrition % =
Overall		
Treatment group (clusters)		
Control group (clusters)		
Differential attrition =		
<b>3/ RISK OF BIAS &lt; JOINERS</b>		
<b>3</b> Is there a risk of bias due to individuals entering clusters (=joiners)? <i>If the study includes joiners in the analytic sample, then the estimate of the effect of the intervention on individual outcomes could be biased if the individuals who entered intervention clusters differ systematically from those who entered comparison clusters. This risk of bias may vary across substantive areas and interventions, and based on how long after random assignment the joining occurred.</i>	<input type="checkbox"/> JA (ga naar 5)	<input type="checkbox"/> NEE ( ga naar 4) <input type="checkbox"/> GEEN INFO ( ga naar 4)
<b>4/ RISK OF BIAS &lt; NON-RESPONSE OF INDIVIDUALS</b>		
<b>4</b> Is there a risk of bias due to non-response of individuals? <i>Nonresponse at the individual level for cluster-level assignment studies is the difference between the individuals present in a reference sample and those present in the analytic sample at the time the outcome is assessed. The reference sample—the benchmark sample from which nonresponse is measured—can differ depending on the risk of bias associated with joiners.</i> <i>Zie bijlage voor een overzicht.</i>	<input type="checkbox"/> JA (ga naar 5) <input type="checkbox"/> GEEN INFO (ga naar 5)	<input type="checkbox"/> NEE ( ga naar 6A)
Berekening		
	N =	Attrition % =
Overall		
Treatment group (individuals)		



Control group (individuals)			
Differential attrition =			

**5/ BASELINE EQUIVALENCE (NIET TE ONDERZOEKEN BIJ CR STUDIE MET LAGE ATTRITIE BIJ CLUSTERS EN BIJ INDIVIDUEN)**

**5** Does the study establish equivalence of individuals at baseline for groups in the analytic sample?  
*BE wordt berekend op basis van Hedges' g.  
 Voor cluster-QE's: indien we geen data vinden voor de analytische sample, maar enkel voor de reference sample, dan zien we dit door de vingers indien er ook onder voorzichtige grenswaarde lage attritie is op vlak van leerlingen (ga hiervoor naar 3) én indien er statistische adjustment gebeurde.  
 Regardless of the level of analysis, this baseline equivalence requirement must be satisfied using individual-level standard deviations. Means calculated using either cluster- or individual-level data are acceptable as long as the weighting is consistent with the weighting used in the analysis. In general, a required statistical adjustment must be made using individual-level data, such that it accounts for the individual-level correlation between the baseline measure and the outcome.*

JA (ga naar 6A)  
 $0 \leq |\text{Baseline ES}| \leq 0.05$   
 OF  
 $0.05 < |\text{Baseline ES}| \leq 0.25$  with acceptable statistic adjustment

NEEN (STOP)  
 $|\text{Baseline ES}| > 0.25$   
 OF  
 $0.05 < |\text{Baseline ES}| \leq 0.25$  without acceptable statistic adjustment

**Berekening**

Outcome measure	Treatment			Control		
	N	M	SD	N	M	SD

**LET OP**

Outcome measures op basis waarvan de TG & CG GEEN baseline equivalentie vertonen, markeer je rood. De effecten gemeten op basis van deze measures (of in deze domeinen) nemen we niet mee in onze bespreking en dus ook niet bij vraag 5)

**6/ OUTCOME MEASURES**

**6** Voldoet minstens één van de outcome measures aan volgende vier voorwaarden:

**6A** The outcome measure demonstrates **face validity**

*An outcome measure must capture what it claims to measure to appropriately label the intervention's effect on an outcome of interest.*

*Face validity:* meet de OM wat hij moet meten? JA indien

- Standardized test\*
- Duidelijke beschrijving van instrument

**6 B** The outcome measure demonstrates **reliability**

*The reliability of a measure captures whether it would yield similar scores from different administrations and whether the observed scores are free of measurement error.*

*Reliability:* is de OM betrouwbaar en consistent? JA indien

- Standardized test\*
- Internal consistency (Cronbach's alpha > 0.50 OF omega > 0.70)
- temporal stability en test-retest stability > 0.40

JA (ga naar DEEL 2)  
 NEEN (STOP)

☐ Inter-rater reliability > 0.50 (%agreement, correlation, kappa)

**6 C The outcome measure is not overaligned with the intervention**

An outcome measure is overaligned if it contains content or materials provided to subjects in one condition but not the other. For example, an outcome measure based on an assessment that relied on reading materials or vocabulary words used in the intervention condition but not in the comparison condition likely would be judged to be overaligned.

This rule does not apply when material covered by an outcome measure must be explicitly taught. For example, reciting the alphabet requires being taught the alphabet, but improving reading comprehension does not require focusing on a specific set of reading passages. Put another way, an outcome measure is only overaligned when the content or materials provided to subjects in a single condition might affect scores on the measure through gaming of the outcome measure, familiarity with the format, or other means besides learning educationally relevant material.

**6 D The outcome measure is collected in the same manner for both groups (T&C)**

Bijvoorbeeld: werd de OM op dezelfde manier, op hetzelfde moment, door hetzelfde type assessor afgenomen?

\* Een OM die aan één van volgende voorwaarden voldoet wordt als ‘standardized’ beschouwd:

- a trademark or a reputable testing organization developed the measure
- measures with technical manuals
- well-known measures for the field and those with established scale scores
- state and district achievement tests

Daarnaast gelden bijkomende criteria

- the measure used in the study must be an established scale or subscale, not just a researcher’s selection of items from the test.
- de test moet door de onderzoekers afgenomen zijn zoals de testprocedures het voorschrijven (bijv. in groep of individueel - doelgroep - correct gebruik scoreprocedures)

Overzicht voorwaarden per outcome measure:

**LET OP:**

bij RCT’s met hoge attritie en QE’s kijken we alleen naar de outcome measures op basis waarvan de analytic sample in vraag 3 equivalent bleek!

Measure	Face validity	Reliability	Not overaligned	Same manner

Werkwijze:

- Zet een kruisje in elke kolom indien JA
- Verantwoord jouw keuze in de cel voor ‘Not overaligned’ indien je oordeelt dat deze uitkomstmaat ‘overaligned’ is
- Zet voor de naam van de uitkomstmaat OK indien deze meegenomen mag worden in onze bespreking.

**7 Imputed data (noteren maar niet mee te nemen in score CA – wel in het achterhoofd houden bij interpretatie analyses)**

7 A Gebruikten de onderzoekers imputed data in hun analyses?	<input type="checkbox"/> JA (ga naar 7 B)	<input type="checkbox"/> NEEN  <input type="checkbox"/> GEEN INFO
7 B Indien de onderzoekers imputed data gebruikten, was dit bij een cluster-RCT met lage attritie?	<input type="checkbox"/> JA  (i.e. laag risico op bias mits een aanvaardbare statistische methode)	<input type="checkbox"/> NEEN  (risico op bias)  <input type="checkbox"/> GEEN INFO

**DEEL 2 Gewenste kwaliteitsindicatoren**

**1/ Sample size**

**1** Is de analytic sample groter dan 250 deelnemers?  
Analytic sample =.....

JA

NEEN  
 GEEN INFO

**2/ Aandacht voor follow-up**

**2** Maten de auteurs het effect van de interventie nog eens (minimum) 1 maand na de post-test?

JA

NEEN

**3/ Implementation fidelity**

<p><b>3</b> Werd de implementation fidelity nagegaan en gerapporteerd?</p>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
<p><b>4/ Ceiling and/or floor effects</b></p>		
<p><b>4</b> Zijn er geen ceiling of floor effecten bij de pre- en posttest?  Bekijk hiervoor de rapportage van de auteurs of onderzoek de distributie van data (vb. boxplot, gegevens ivm range of normale distributie).  <i>Graham &amp; Santangelo; Hebert et al 2018: Mean pretest &amp; posttest scores for students in the intervention and comparison conditions were more than one standard deviation away from the lowest and highest score of the scale for each pretest measure used in our analysis. True random experiments that did not include a pre-test were credited as having met this criterion.</i></p>	<input type="checkbox"/> JA (ER ZIJN DUIDELIJK GEEN CEILING OF FLOOR EFFECTEN)	<input type="checkbox"/> NEEN (ER ZIJN CEILING OF FLOOR EFFECTEN)  <input type="checkbox"/> GEEN INFO (IK KAN DIT NIET ACHTERHALEN)
<p><b>5/ Blinding</b></p>		
<p><b>5</b> Werden de pre- en post-testen blind afgenomen door de assessors?  <i>Indien de leerkrachten zelf de test afnemen bij hun leerlingen, coderen we dit als een niet-blinde testafname. Indien sommige testen blind werden afgenomen en andere niet, scoor dan JA.</i></p>	<input type="checkbox"/> JA (BLIND AFGENOMEN)	<input type="checkbox"/> NEEN <input type="checkbox"/> GEEN INFO
<p><b>6/ At-risk status</b></p>		
<p><b>6</b> Werd duidelijk beschreven op basis van welke definitie of measures deelnemers tot de at-risk groep gerekend werden?</p>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN
<p><b>7/ Beschrijving controlegroep</b></p>		
<p><b>7</b> Werd duidelijk beschreven welk programma de controlegroep als alternatief voor de interventie kreeg?</p>	<input type="checkbox"/> JA	<input type="checkbox"/> NEEN

## Critical appraisal – Effectstudie met Individual-level assignment

<b>Titel</b>	
<b>Auteur(s)</b>	
<b>Jaartal</b>	

<b>SCORE</b>	<b>DEEL 1</b>
<b>DEEL 1:</b> 0 / -2	<b>1</b> = RCT's met lage attritie
<b>DEEL 2:</b> /7	<b>-2</b> = Alle andere designs
<b>Totaalscore</b> /7	<b>DEEL 2</b> Elke JA is een punt. 2 x GEEN INFO is een punt.

<b>BESLISSING</b>	
- $\geq 4$ = HK	<input type="checkbox"/> <b>HK= HOGE KWALITEIT</b>
- $< 4$ = MK	<input type="checkbox"/> <b>MK= MIDDELMATIGE KWALITEIT</b>
- <b>STOP in DEEL 1 = LK</b>	<input type="checkbox"/> <b>UIT= LAGE KWALITEIT</b>

### DEEL 1 Vereiste kwaliteitsindicatoren

#### 0/ CONFOUNDING FACTORS

**0A** Does the intervention or comparison group contain a **single unit** (n=1)?

*The intervention or comparison group contains a single study unit—such as a teacher, classroom, or school—and that unit is not present in the other condition.*

*Bijvoorbeeld: two schools are randomly assigned, one to each condition; a study has two intervention classrooms and two comparison classrooms, but both intervention classrooms had the same teacher, who had no interaction with the comparison classrooms.*

JA (STOP)

NEEN (ga naar 0B)

**0B** Do the characteristics of the subjects in the intervention or comparison group differ systematically, **with no overlap**, in ways that are associated with the outcomes?

*Bijvoorbeeld: Teacher education level: a small group of teachers in a master's program implements the intervention, whereas students in the comparison group are taught by teachers with bachelor's degrees.*

*Different time period (bijv. successive-cohort designs)  
Student level: Intervention students are all in grade 8, and comparison students are all in grade 7.*

*Student characteristics: Intervention students are all English learners, and no comparison students are.*

JA (STOP)

NEEN (ga naar 1A)

#### 1/ RANDOM ASSIGNMENT

<b>1A</b> Were the participants randomly assigned to treatment and control groups?	<input type="checkbox"/> JA (ga naar 1B)	<input type="checkbox"/> NEEN (ga naar 3)
<b>1B</b> Are the validity or the analysis of the random assignment process compromised? <i>Bijvoorbeeld:</i> - Een deelnemer wordt na RA nog van groep veranderd - RA to a group with different probabilities	<input type="checkbox"/> JA (ga naar 3)	<input type="checkbox"/> NEEN (ga naar 2)

**2/ ATTRITION (ENKEL TE CONTROLEREN BIJ RCT)**

<b>2</b> Is the combination of overall and differential attrition acceptable?  <i>Indien er in de RCT meer dan twee groepen opgenomen zijn, vergelijk dan de twee groepen met het grootste verschil in attritie.</i> <i>We kiezen voor de voorzichtige grenswaarde, tenzij de auteurs aangeven dat de uitval niet te maken had met de interventie (vb. ziekte van het kind, personeelsproblemen).</i>	<input type="checkbox"/> JA (ga naar 4) Zie tabel in bijlage	<input type="checkbox"/> NEEN / GEEN INFO (ga naar 3) Zie tabel in bijlage
--	---	---

Berekening

	N =	Attrition % =
Overall		
Treatment group		
Control group		
Differential attrition =		

**3/ BASELINE EQUIVALENCE (NIET VOOR RCT MET LAGE ATTRITIE)**

<b>3</b> Is equivalence established for the groups in the <u>analytic sample</u> ? <i>For high attrition RCTs &amp; QE's, equivalence of the intervention and comparison groups on specified characteristics measured at baseline—that is, prior to the introduction of the intervention—must be assessed for the analytic sample, the subjects from the intervention and comparison groups used to estimate findings.</i> <i>Voor QE's: indien we geen data vinden voor de analytische sample, maar enkel voor de reference sample, dan mag het artikel toch door indien er ook onder voorzichtige grenswaarde lage attritie is (ga hiervoor naar 3) én indien er statistische adjustment gebeurde.</i>	<input type="checkbox"/> JA (ga naar 4) $0 \leq  \text{Baseline ES}  \leq 0.05$ OF $0.05 <  \text{Baseline ES}  \leq 0.25$ with acceptable statistic adjustment	<input type="checkbox"/> NEEN (STOP & UIT) $ \text{Baseline ES}  > 0.25$ OF $0.05 <  \text{Baseline ES}  \leq 0.25$ without acceptable statistic adjustment
--	--	--

Berekening

Outcome measure	Treatment			Control		
	N	M	SD	N	M	SD


**LET OP**

Outcome measures op basis waarvan de TG & CG GEEN baseline equivalentie vertonen, markeer je rood. De effecten gemeten op basis van deze measures (of in deze domeinen) nemen we niet mee in onze bespreking en dus ook niet bij vraag 4)

**4/ OUTCOME MEASURES**

**4 Voldoet minstens één van de outcome measures aan volgende vier voorwaarden:**

**4A The outcome measure demonstrates face validity**  
*An outcome measure must capture what it claims to measure to appropriately label the intervention's effect on an outcome of interest.*

*Face validity:* meet de OM wat hij moet meten? JA indien

- Standardized test\*
- Duidelijke beschrijving van instrument

**4B The outcome measure demonstrates reliability**  
*The reliability of a measure captures whether it would yield similar scores from different administrations and whether the observed scores are free of measurement error.*

*Reliability:* is de OM betrouwbaar en consistent? JA indien

- Standardized test\*
- Internal consistency (Cronbach's alpha > 0.50 OF omega > 0.70)
- temporal stability en test-retest stability > 0.40
- Inter-rater reliability > 0.50 (%agreement, correlation, kappa)

**4C The outcome measure is not overaligned with the intervention**  
 An outcome measure is overaligned if it contains content or materials provided to subjects in one condition but not the other. For example, an outcome measure based on an assessment that relied on reading materials or vocabulary words used in the intervention condition but not in the comparison condition likely would be overaligned. This rule does not apply when material covered by an outcome measure must be explicitly taught. For example, reciting the alphabet requires being taught the alphabet, but improving reading comprehension does not require focusing on a specific set of reading passages. Put another way, an outcome measure is only overaligned when the content or materials provided to subjects in a single condition might affect scores on the measure through gaming of the outcome measure, familiarity with the format, or other means besides learning educationally relevant material.

**4D The outcome measure is collected in the same manner for both groups (T&C)**  
 Bijvoorbeeld: werd de OM op dezelfde manier, op hetzelfde moment, door hetzelfde type assessor afgenomen?

- \* Een OM die aan één van volgende voorwaarden voldoet wordt als 'standardized' beschouwd:
- a trademark or a reputable testing organization developed the measure
  - measures with technical manuals
  - well-known measures for the field and those with established scale scores

- JA (ga naar DEEL 2)
- NEEN (STOP & UIT)

<ul style="list-style-type: none"> <li>- state and district achievement tests</li> </ul> <p>Daarnaast gelden bijkomende criteria</p> <ul style="list-style-type: none"> <li>- the measure used in the study must be an established scale or subscale, not just a researcher's selection of items from the test.</li> <li>- de test moet door de onderzoekers afgenomen zijn zoals de testprocedures het voorschrijven (bijv. in groep of individueel - doelgroep - correct gebruik scoreprocedures)</li> </ul>	
--	--

Overzicht voorwaarden per outcome measure:

**LET OP:**

bij RCT's met hoge attritie en QE's kijken we alleen naar de outcome measures op basis waarvan de analytic sample in vraag 3 equivalent bleek!

Measure	Face validity	Reliability	Not overaligned	Same manner

Werkwijze

- Zet een kruisje in elke kolom indien JA
- Verantwoord jouw keuze in de cel voor 'Not overaligned' indien je oordeelt dat deze uitkomstmaat 'overaligned' is
- Zet voor de naam van de uitkomstmaat OK indien deze meegenomen mag worden in onze bespreking.

**7 Imputed data (noteren maar niet mee te nemen in score CA – wel in het achterhoofd houden bij interpretatie analyses)**

7 A Gebruikten de onderzoekers imputed data in hun analyses?	<input type="checkbox"/> JA (ga naar 7 B)	<input type="checkbox"/> NEEN  <input type="checkbox"/> GEEN INFO
--	---	---



<p>7 B Indien de onderzoekers imputed data gebruikten, was dit bij een cluster-RCT met lage attritie?</p>	<p><input type="checkbox"/> JA  (i.e. laag risico op bias mits een aanvaardbare statistische methode)</p>	<p><input type="checkbox"/> NEEN  (risico op bias)</p> <p><input type="checkbox"/> GEEN INFO</p>
---	---	--

<h2 style="text-align: center;">DEEL 2 Gewenste kwaliteitsindicatoren</h2>		
<p><b>1/ Sample size</b></p>		
<p><b>1</b> Is de analytic sample groter dan 250 deelnemers? Analytic sample =.....</p>	<p><input type="checkbox"/> JA</p>	<p><input type="checkbox"/> NEEN</p>
<p><b>2/ Aandacht voor follow-up</b></p>		
<p><b>2</b> Maten de auteurs het effect van de interventie nog eens (minimum) 1 maand na de post-test?</p>	<p><input type="checkbox"/> JA</p>	<p><input type="checkbox"/> NEEN</p>
<p><b>3/ Implementation fidelity</b></p>		
<p><b>3</b> Werd de implementation fidelity nagegaan en gerapporteerd?</p>	<p><input type="checkbox"/> JA</p>	<p><input type="checkbox"/> NEEN</p>
<p><b>4/ Ceiling and/or floor effects</b></p>		
<p><b>4</b> Zijn er geen ceiling of floor effecten bij de pre- en posttest? Bekijk hiervoor de rapportage van de auteurs of onderzoek de distributie van data (vb. boxplot, gegevens ivm range of normale distributie). <i>Graham &amp; Santangelo; Hebert et al 2018: Mean pretest &amp; posttest scores for students in the intervention and comparison conditions were more than one standard deviation away from the lowest and highest score of the scale for each pretest measure used in our analysis. True</i></p>	<p><input type="checkbox"/> JA (ER ZIJN DUIDELIJK GEEN CEILING OF FLOOR EFFECTEN)</p>	<p><input type="checkbox"/> NEEN (ER ZIJN CEILING OF FLOOR EFFECTEN)</p> <p><input type="checkbox"/> GEEN INFO (IK KAN DIT NIET ACHTERHALEN)</p>

<p><i>random experiments that did not include a pre-test were credited as having met this criterion.</i></p>		
<p><b>5/ Blinding</b></p>		
<p><b>5</b> Werden de pre- en post-testen blind afgenomen door de assessors?  <i>Indien de leerkrachten zelf de test afnemen bij hun leerlingen, coderen we dit als een niet-blinde testafname. Indien sommige testen blind werden afgenomen en andere niet, scoor dan JA.</i></p>	<p><input type="checkbox"/> JA  (BLIND AFGENOMEN)</p>	<p><input type="checkbox"/> NEEN  <input type="checkbox"/> GEEN INFO</p>
<p><b>6/ At-risk status</b></p>		
<p><b>6</b> Werd duidelijk beschreven op basis van welke definitie of measures deelnemers tot de at-risk groep gerekend werden?</p>	<p><input type="checkbox"/> JA</p>	<p><input type="checkbox"/> NEEN</p>
<p><b>7/ Beschrijving controlegroep</b></p>		
<p><b>7</b> Werd duidelijk beschreven welk programma de controlegroep als alternatief voor de interventie kreeg?</p>	<p><input type="checkbox"/> JA</p>	<p><input type="checkbox"/> NEEN</p>