

Gerandomiseerd onderzoek met controlegroep: op zoek naar effecten

Artur Jaschke

Om erachter te komen welke effecten een bepaalde aanpak of interventie sorteert, geldt gerandomiseerd experimenteel onderzoek als de gouden standaard. Artur Jaschke belicht in dit artikel de kenmerken, eisen, kracht en beperkingen van deze onderzoeksmethode.

Er is veel onderzoek gedaan naar de effecten van muziek en in het bijzonder muziekonderwijs op het menselijke brein (Schlaug, 2015; Zatorre & Belin, 2001; Rose, 2009). Vaak wordt onderzocht hoe muziekonderwijs aangeboden wordt, hoe dit soort onderwijs vorm gegeven moet worden en hoe kinderen ermee omgaan. Het gaat hier om zogeheten *near transfer*-effecten, bijvoorbeeld als een kind elke dag piano speelt, is de kans groot dat het beter wordt in piano-spelen. Om dit soort effecten naar een hoger plan, *far transfer*, te tillen halen onderzoekers vaak de gedrags- en neurowetenschappen erbij. Ze kijken dan of datgene wat een kind leert door kunst- of muziekonderwijs te *vertalen* is naar andere cognitieve taken zoals algemene schoolprestaties (Jaschke, Honing, & Scherder, 2018; Holochwost et al., 2017).

Om deze bredere transfereffecten aan te kunnen tonen zijn methodes als observatie of vragenlijsten vaak onvoldoende. Ook is het lastig om een transfereffect aan te tonen als interventies te kort aangeboden worden of onduidelijk of zelfs helemaal niet worden beschreven (Dége, Kubicek, & Schwarzer, 2011). Hiervoor is een experimentele onderzoeksopzet nodig met grote groepen proefpersonen die blind, en als het even kan 'dubbelblind' verdeeld worden over interventie- en controlegroep, en gerandomiseerd en longitudinaal getest worden met een aantal cognitieve taken dan wel hersenscans. Wees gerust als u de vorige zin niet direct hebt begrepen. In dit artikel zal ik uitleggen wat zo'n 'longitudinal single blinded randomised controlled trial' precies behelst.

Onderzoeksvraag en protocol

Elk onderzoek begint met een vraag. Zo ook een *randomised controlled trial* (RCT). Bij een RCT vertaalt de onderzoeker die vraag naar een onderzoek-protocol: een nauwkeurige beschrijving van de onderzoeksmethode, het benodigd aantal en soort proefpersonen en meetinstrumenten. Dit protocol moet de onderzoeker ter beoordeling en goedkeuring voorleggen aan een medisch-ethische commissie of een commissie wetenschap en ethiek voordat het onderzoek mag beginnen. Academisch medische centra en universiteiten hebben dergelijke commissies, maar ze komen ook steeds vaker in het hoger beroepsonderwijs voor. Zonder goedkeuring mag er geen grootschalig onderzoek uitgevoerd worden met medisch-wetenschappelijke apparatuur zoals (neuro)psychologische testen of hersenscans of, meer in het algemeen, met mensen of dieren.

'Zulke testen hebben toch niets met muziekonderwijs te maken', zult u wellicht denken, maar het tegenovergestelde is waar. Om een effect, in dit geval van muziekonderwijs, aan te kunnen tonen en als werkzaam of zinvol aan te kunnen duiden, los van het effect dat een kind vaardiger wordt in het bespelen van een instrument, zijn inzichten in cognitieve prestaties essentieel.

Kenmerken RCT

Bij een RCT voert de onderzoeker de te testen behandeling, in dit geval muziekonderwijs, uit bij een zogeheten interventiegroep (de groep die muziekonderwijs krijgt) en een of meer controlegroepen die een andere of helemaal geen interventie krijgen.

Een controlegroep is een groep proefpersonen die in zoveel mogelijk opzichten vergelijkbaar is met de interventiegroep, zodat het enige verschil tussen beide groepen de interventie is. In onderzoeken worden proefpersonen met elkaar *gematched*. Dit betekent dat ze een aantal testen moeten doen voordat de interventie begint, de zogenoemde *baselinemeeting*. Door de resultaten naast elkaar te leggen kan de onderzoeker kijken of de proefpersonen inderdaad vergelijkbaar zijn. Bij deze aanpak vallen vaak al de eerste proefpersonen af. Als ze wel mee zouden doen, zouden de resultaten van het onderzoek namelijk niet betrouwbaar zijn.

Om te voorkomen dat er bij het indelen van de groepen een verschil ontstaat dat de kans op werking van de interventie kan beïnvloeden, schakelt de onderzoeker vaak een derde partij in die de proefpersonen verdeelt over interventie- of controlegroep: het randomiseren. Dit wordt gedaan om te voorkomen dat de onderzoeker onbewust zou sturen, bijvoorbeeld door de proefpersonen die een beetje beter op de *baseline tests* gepresteerd hebben, toe te wijzen aan de interventiegroep. Dergelijke selectie-effecten kunnen zeer subtiel zijn. Bij het werken met proefdieren zien we dat zelfs het met de hand vangen van 'willekeurige' proefdieren uit een kooi, bijvoorbeeld al een duidelijk selectie-effect kan geven, doordat tamme, dikke of trage dieren meer kans hebben te worden gepakt.

Het zou natuurlijk wenselijk zijn dat de uitvoerende onderzoeker noch de proefpersoon weet wie welke interventie krijgt. Dat is het zogenoemde *dubbel-blind design*, zoals bij een onderzoek met een medicijn en een placebo de proefpersonen niet weten welke pil ze slikken. Dit is helaas niet altijd mogelijk en in het geval van onderzoek naar (muziek)onderwijs zelfs geheel onmogelijk. Daarom wordt in dit geval een *single blinded* opzet gebruikt, waar proefpersonen en onderzoeker weten welke kinderen welke interventie krijgen.

Om een gerandomiseerd onderzoek met controlegroepen nog betrouwbaarder te maken lassen onderzoekers vaak in plaats van twee meetmomenten, één voor en één na de interventie, meer momenten in. Het gaat hierbij om van te voren geprotocolleerde testen. Door dezelfde testen over een langere periode, gemiddeld anderhalf tot drie jaar, af te nemen bij proefpersonen, valt meer een lijn van prestatie vast te stellen en zijn effecten die pas na één of twee jaar eventueel optreden, veel nauwkeuriger te beschrijven. Binnen het muziekonderwijs is bijvoorbeeld gebleken dat kinderen in de muziekgroep pas na anderhalf jaar muziekonderwijs vooruit gingen op cognitieve prestaties en executieve functies als planning, werkgeheugen en aandacht (Jaschke et al., 2018).

Experimenteel proces

Omdat het uitvoeren van experimenten tijd, materiaal en menskracht in beslag neemt, staan onderzoekers voor een dilemma: voor een betrouwbaar onderzoeksresultaat zijn een groot aantal proefpersonen en goede materialen en methodes noodzakelijk, maar tegelijkertijd mag het beslag op tijd en proefpersonen niet groter zijn dan strikt noodzakelijk. Door het onderzoek goed te plannen en alle inclusie- en exclusiecriteria ruim van te voren te bepalen ontstaat een gedegen onderzoeksprotocol waarmee de onderzoeker nauwkeurig het verband tussen onafhankelijke variabelen ('knoppen' waaraan gedraaid kan worden) en afhankelijke variabelen (procesuitkomsten) kan bepalen. Voorbeelden van criteria om proefpersonen uit te sluiten (exclusie) kunnen zijn: slecht gehoor, motorische achterstand en andere fysiologische onderdelen die het onmogelijk maken om deel te nemen aan de interventie of de testen. Inclusiecriteria zijn leeftijd, sekse en/of socio-economische achtergrond. De criteria verschillen per onderzoek en kunnen lichamelijke, geestelijke, maatschappelijke en culturele factoren betreffen.

Kort samengevat kan een experimenteel proces er als volgt uit zien:

- *Vorbereitung*
In de voorbereidingsfase denkt de onderzoeker onder meer na over het doel van het experiment, de verschillende parameters en hoe deze te meten. Dit gebeurt vanuit de hypothese oftewel de onderzoeksvraag.
- *Opzetten experiment*
In deze fase legt de onderzoeker het protocol voor aan een ethische of wetenschappelijke commissie, gaat hij de proefpersonen benaderen, inclusiecriteria toetsen (met vragenlijsten en statistische analyses) en alle materialen klaarmaken voor de uitvoer van het onderzoek.
- *Uitvoeren experiment*
Op basis van de in de vorige fase bepaalde volgorde voert de onderzoeker de testen uit en stelt hij de resultaten vast.
- *Analyse resultaten*
De onderzoeker analyseert de verkregen resultaten statistisch en presenteert deze doorgaans in een inzichtelijke grafiek.
- *Conclusies*
De onderzoeker valideert indien mogelijk de resultaten, beschrijft en publiceert ze om deze breed te delen met de wetenschappelijke gemeenschap.

Cognitieve vaardigheden

Om de algemene effecten van muziekonderwijs op gedrag en hersengroei te kunnen bepalen zet de onderzoeker in samenwerking met een neuropsycholoog en neurowetenschapper een aantal testen bij elkaar die gericht cognitieve vaardigheden zoals plannen, werkgeheugen of aandacht kunnen meten. Het gaat hierbij niet om een *near transfer*, maar juist naar bredere cognitieve functies die niet alleen bij muziek, maar ook bij bijvoorbeeld rekenen en taal een rol spelen.

In een recent onderzoek werden 174 kinderen met elkaar vergeleken op executieve functies, verbale intelligentie en prestaties op Cito-volgtolsten. Het onderzoek was opgezet als een *single blinded longitudinal randomised controlled trial*, waarbij de proefpersonen (leerlingen) werden verdeeld over vier groepen: 1) muzikles op school; 2) muzikles op school en thuis; 3) beeldende kunst les op school en 4) geen toegevoegd kunst- en muziekonderwijs.

De kinderen waren gerandomiseerd toegewezen aan één van deze vier groepen (Jaschke et al., 2018). Zij werden tweeënehalf jaar gevolgd en elke zes maanden getest met een combinatie (testbatterij) van neuropsychologische testen en taken die verschillende cognitieve of lichamelijke functies maten. Neuropsychologische testen meten vaak cognitieve prestaties en veranderingen en kunnen ook zicht geven op groei, achteruitgang en ontwikkeling van de hersenen. Uit dit onderzoek kwam naar voren dat kinderen in de muzikgroepen, ongeacht enkel thuis of gecombineerd met muzikles op school, beter presteerden op de taken die executieve functies vereisten. Kinderen in de beeldende kunstgroep presteerden beter op taken die om ruimtelijk inzicht vroegen.

Deze resultaten wijzen erop dat langdurig muziekonderwijs een positieve invloed kan hebben op cognitieve vaardigheden zoals planning, remming en aandacht op kunnen treden. Deze resultaten ondersteunen een mogelijke *far transfer*, naar rekenen en taal, omdat de gemeten cognitieve functies ook onderliggend zijn voor rekenen en taal. Er zijn vergelijkbare tot dezelfde hersengebieden en netwerken nodig voor het maken van muziek en voor bijvoorbeeld het oplossen van een gecompliceerde rekensom. Dat wil niet zeggen dat wij slimmer worden door muziek te spelen. Het Mozart-effect – of het Beyoncé-effect, Gustav Mahler-effect, John Coltrane-effect of Hazes-effect – bestaat niet. Maar muziek stimuleert wel het hele brein, waaronder executieve functies, geheugen en emoties. Door deze grootschalige stimulering van de hersenen worden nieuwe verbindingen gelegd en bestaande verbindingen versterkt. Zo krijgen wij door muziek te maken een beter gestimuleerd brein ‘cadeau’.

Een methodologische beperking in deze studie is het meten van Cito-scores. Hoewel de Cito-volgtolsten leerlingen op taken als kritisch luisteren of schrijven beoordelen, hebben de onderzoekers deze scores niet

vergeleken met andere testen, zoals testen die bijvoorbeeld fonologisch bewustzijn meten. Vergelijking van de Cito-scores, die een gegeneraliseerd speciaal model zijn voor Nederlandse kinderen, met meer internationaal gestandaardiseerde taalfunctietesten (bijvoorbeeld phonological assessment battery; Gallagher & Frederickson, 1995), zou meer variatie in de executieve functies kunnen opleveren en dus andere resultaten. Maar zo'n vergelijking zou de testtijd met ruim drie uur verhogen, rekening houdend met alle academische vaardigheden en alle subcomponenten (zoals fonologisch bewustzijn, semantiek en lexemen in taal en logica, abstract denken en rekenen in wiskunde). Langere testtijden kunnen invloed hebben op de concentratie van deelnemers. Bovendien was deze studie alleen gericht op verbaal IQ, en niet op een volledige IQ-schaal, wat mogelijk een ander resultaat op intelligentiemaatregelen kan aantonen (Schellenberg, 2006; Dumont, Syurina, Feron, & Van Hooren, 2017).

Discussie

Experimenteel onderzoek als RTC wordt gebruikt om de effecten van iets op iets of iemand te meten. Alleen door verdere ontwikkeling, trial-and-error, komen onderzoekers erachter hoe bijvoorbeeld onderwijs gegeven moet worden. Experimenteel onderzoek zoekt niet naar het hoe (dat past beter in een kwalitatieve methode), maar meet louter de effecten van het hoe op het wat.

Ook al geldt gerandomiseerd onderzoek nog steeds als een van de beste methodes om duidelijke effecten aan te kunnen tonen, dit soort onderzoek kent ook zijn zwaktes en beperkingen. Vaak vallen er veel proefpersonen af, omdat er bijwerkingen zijn of, in de context van onderwijs, kinderen afhankelijk zijn van hun thuissituatie of ziek kunnen worden. Het kan zijn dat ze niet alle meetmomenten bijgewoond hebben of dat de testresultaten niet bruikbaar zijn omdat proefpersonen deze niet bij elke meting af hebben gemaakt. Bij het analyseren van de resultaten worden deze proefpersonen niet meegeteld, wat mogelijk een kleiner effect oplevert. Dat is eigenlijk jammer, omdat men nooit meer te weten komt of deze proefpersonen baat hadden bij de interventie: ze worden domweg weggelaten en er wordt nooit meer naar gekeken.

Daarnaast zijn er ethische aspecten. Zo liet het onderzoek van Jaschke en collega's (2018) zien dat kinderen in de muziekgroepen betere Cito-scores kregen en hun executieve functies beter leerden te gebruiken door het spelen van muziek. De controlegroep heeft deze ontwikkeling in Cito-scores en executieve functies niet op dezelfde schaal kunnen meemaken. Zulke bezwaren zijn in principe te ondervangen door na afloop van een RCT de controlegroep de interventie alsnog aan te bieden. Toch zouden de effecten, vooral bij kinderen die een heel snelle en dynamische hersenontwikkeling kennen, misschien kleiner zijn, omdat hun brein al wat stappen verder is en

zo het effect op hersengroei en gedrag minder groot zou kunnen uitvallen. Al weten we inmiddels dat het brein plastisch is en gedurende het hele leven kan veranderen en nieuwe cellen aanmaken, als wij het maar blijven stimuleren en uitdagen (Rodrigues, Loureiro, & Caramelli, 2010).

Een andere beperking is dat er veel achtergrondkennis over de te onderzoeken groep, interventie en factoren nodig is, zoals socio-economische achtergrond van de proefpersonen, om de resultaten goed te kunnen interpreteren en te vertalen naar de dagelijkse praktijk. Hierbij komt vaak een analyse kijken die het minimum aantal proefpersonen berekent aan de hand van interventie-effecten en gemiddelde resultaten die het dichtst aansluiten bij de eigen onderzoeksvraag. Hierbij mag de relevantie van de behaalde onderzoeksresultaten niet vergeten worden. De onderzoeker moet deze adequaat interpreteren en de relevantie van ook maar de kleinste verbeteringen of verslechtingen juist interpreteren en implementeren in vervolgonderzoek. Aanvullend moet de onderzoeker verwachte verschillen binnen de groepen begrijpen en analyseren, om überhaupt een resultaat te kunnen zien in de te onderzoeken groep. Dat kost tijd en geld.

Zou bij het behalen van negatieve resultaten (dus geen gevonden effecten), muziek misschien helemaal verdwijnen uit het onderwijs, de kliniek of ons leven? Het lijkt me niet dat uitkomsten zo'n vergaande invloed moeten hebben op wel of geen kunst en cultuur. Muziek en in bredere zin cultuur, is een onderdeel van onze samenleving. Het is niet meer weg te denken en zou in welke vorm dan ook moeten blijven bestaan.

Wereldwijd gebruikt men experimenteel onderzoek om effecten van kunst- en cultuuronderwijs te meten. Onderzoeksgroepen in elk Europees land, in de VS en Canada, maar ook in Azië en Oceanië, zien het belang van kunst- en cultuuronderwijs en proberen dit experimenteel aan te tonen. Het is uitermate belangrijk wat precies de onderzoeksvraag, de context en de aanpak van het onderzoek is. Willen wij zien hoe wij beter onderwijs kunnen geven en wat er in het onderwijs nodig is, dan vormt een kwalitatieve onderzoeksopzet een goed begin, waarbij in een later stadium de implementatie met een experimenteel onderzoek getoetst kan worden. Bij een complex thema als onderwijs is altijd een combinatie van verschillende methodes en aanpakken nodig.

Gezien de positieve effecten die steeds meer onderzoek laten zien, is het niet ons doel als wetenschappers om muziek het stempel *ja of nee* te geven, maar om te begrijpen wat muziek nog meer voor ons kan doen, naast het reguleren van onze emoties, ons beter onze executieve functies te leren gebruiken, samen iets te doen waar wij plezier aan beleven en uiteraard onze hersen en netwerken verder te versterken. Het is nooit te laat om te beginnen met een muziekinstrument! Elke muziekstijl hierin is van belang als de hersenen maar uitgedaagd worden met iets nieuws, dag in, dag uit.

Artur Jaschke is onderzoeker bij de afdeling Klinische neuropsychologie van de Vrije Universiteit en de afdeling Neonatologie aan het Universitaire Medisch Centrum Groningen. Hij doet onder meer onderzoek naar de relatie tussen muziek en breinripping. Hij is daarnaast lector Muziektherapeutische Interventies aan ArtEZ Conservatorium Enschede.
E a.c.jaschke@vu.nl

Literatuur

Degé, F., Kubicek, C., & Schwarzer, G. (2011). Music lessons and intelligence: a relation mediated by executive functions. *Music Perception*, 29(2), 195-201.

Dumont, E., Syurina, E. V., Feron, F. J. M., & Hooren, S. van. (2017). Music interventions and child development: a critical review and further directions. *Frontiers in Psychology*, 8, [1694]. DOI: 10.3389/fpsyg.2017.01694

Gallagher, A., & Frederickson, N. (1995). The Phonological Assessment Battery (PhAB): an initial assessment of its theoretical and practical utility. *Educational and Child Psychology*, 12(1), 53-67.

Holochwost, S., Propper, C., Wolf, D., Willoughby, M., Fischer, K., Kolacz, J., et al. (2017). Music education, academic achievement, and executive functions. *Psychology of Aesthetics, Creativity, and the Arts*, 11(2), 147-166.

Jaschke A. C., Honing, H., & Scherder, E. J. A. (2018). Longitudinal analysis of music education on executive functions in primary school children. *Frontiers in Neuroscience*, 12, [103]. DOI: 10.3389/fnins.2018.00103

Rose, F. C. (2009). *Neurology of Music*. London: Imperial College Press.

Rodrigues, A. C., Loureiro, M. A., & Caramelli, P. (2013). Long-term musical training May improve different forms of visual attention ability. *Brain and Cognition*, 82(3), 229-235.

Schellenberg, E. G. (2006). Long-term positive associations between music lessons and IQ. *Journal of Educational Psychology*, 98(2), 457-468.

Schlaug, G. (2015). Musicians and music making as a model for the study of brain plasticity. In E. Altenmüller & F. Boller (Eds.), *Music, Neurology, and the Brain* (pp. 37-49). Amsterdam/New York, NY: Elsevier.

Zatorre, R. J & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, 11(10), 946-953.