

Empirische Sonderpädagogik, 2015, Nr. 3, S. 269-279
ISSN 1869-4845 (Print) · ISSN 1869-4934 (Internet)

Kritische Aspekte bei Blickbewegungsanalysen am Beispiel der Evaluierung von Trainingseffekten im schlussfolgernden Denken bei Risikogruppen

Marco G.P. Hessels¹ & Susanne Schwab²

¹ *Universität Genf*

² *Universität Bielefeld*

Zusammenfassung

Im Rahmen des vorliegenden Beitrags werden Problematiken von Blickbewegungsanalysen anhand von Studien, welche Trainingseffekte im schlussfolgerndem Denken von Kindern mit und ohne Lernschwierigkeiten sowie Jugendlichen mit (leichter) geistiger Behinderung evaluieren, kritisch aufgezeigt. Das Trainieren von Regeln und Lösungsprozessen ist Bestandteil von dynamischen Tests, welche z.B. die Fähigkeit zum schlussfolgernden Denken messen. Die Evaluation der Trainingseffekte kann u.a. mit Blickbewegungsanalysen vorgenommen werden, allerdings ergeben sich dabei methodische Probleme. Diese hängen unter anderem mit dem häufigeren Verlust von Blickbewegungsaufzeichnungen, infolge von ungünstigen Testbedingungen bei der Durchführung von Studien mit Kindern oder Personen mit Behinderungen, zusammen.

Schlüsselwörter: Blickbewegungen, Dynamisches Testen, Genauigkeit, Lernschwierigkeiten

Methodological concerns regarding eye movement analysis illustrated with the assessment of training effects in inductive reasoning in at risk groups

Abstract

In this article several difficulties are illustrated that may arise when using eye movement equipment to analyze changes in problem solving behavior of young children and adolescents with learning difficulties and/or intellectual disabilities as a result of training. Training of rules and procedures in analogical reasoning tasks is used, e.g., in dynamic measures of reasoning capacity. Several eye-movement studies investigating whether problem solving processes indeed change as a result of training are presented to illustrate some methodological concerns. These are related to data loss due to imperfect conditions when assessing young children or adolescents with disabilities, but also to the precision and validity of the eye-tracking data obtained.

Keywords: eye movement, precision, dynamic assessment, learning difficulties

Die Aufzeichnung von Blickbewegungen geriet in den letzten Jahren immer stärker in den Fokus von unterschiedlichsten Forschungen. Videobasierte Eye-Tracking-Systeme erlauben eine relativ einfache Anwendung (im Vergleich zu anderen Methoden wie z.B. EEG-Analysen) in vielen unterschiedlichen Settings und demzufolge werden diese auch immer häufiger angewendet (Holmqvist et al., 2011). Viele Studien nutzen Blickbewegungsanalysen, um tiefere Einblicke darüber zu bekommen, wie Personen Aufgaben lösen bzw. Informationen verarbeiten (z.B. Zhao, Schnotz, Wagner & Gaschler, 2014). Rayner (1998) setzt voraus, dass ein starker Zusammenhang zwischen der Aufmerksamkeit und den Blickbewegungen in komplexen Informationsverarbeitungsaufgaben existiert. Eye-Tracker liefern die Informationen darüber, zu welchem Zeitpunkt (Zeitstempel) sich der Blick wo befindet (Blickpositionen). Maße, welche mit der Tiefe der kognitiven Informationsverarbeitung und der Aufmerksamkeitsfokussierung (welche Information wird beachtet und was wird ignoriert) im Zusammenhang stehen, sind die Anzahl und die Dauer von Blickfixierungen im Bezug auf sogenannte „Areas of Interest“ (AOIs, genaue Teilbereiche des Sichtfeldes, auf die sich die jeweilige Forschung konzentriert). Die jeweils relevanten Maße, z.B. Dauer der Verweilung in den AOIs oder das Springen des Blicks von einer AOI zur anderen, werden dabei über alle Items oder über einzelne Itemsblöcke aggregiert (siehe dazu z.B. auch Gaschler, Marewski & Frensch, in Druck; Zhao et al., 2014) und können in gängigen Programmen (wie z.B. SPSS, R oder Microsoft-Excel) ausgewertet werden.

Technisch ist es mittlerweile möglich, Eye-Tracking-Systeme in Schulklassen zu verwenden, da kein schwierig transportierbares Equipment benötigt wird und auch auf Kopfstützen verzichtet werden kann. Wenngleich nun also die technische Umsetzung und die Datenauswertung relativ simpel sind, so können sich bei der Durchführung, der Datenaufzeichnung und den da-

rauf basierenden Interpretationen von Blickbewegungen gravierende Problematiken ergeben, speziell wenn die Testpersonen Kinder oder Personen mit Behinderung (z.B. Lernbehinderung oder geistiger Behinderung) sind. Der vorliegende Beitrag versucht dies am Beispiel von Aufgaben zum schlussfolgernden Denken in Lerntests aufzuzeigen.

Lerntests bei Risikogruppen

Die Verwendung traditioneller Intelligenztests wurde für Kinder mit Lernschwierigkeiten oder (leichter) geistiger Behinderung in Bezug auf die geringe Reliabilität und mangelnde Vorhersagevalidität oftmals kritisiert (z.B. Budoff, 1967; Hessels, 2009; Tiekstra, Hessels & Minnaert, 2009; Wishart & Duffy, 1990). Traditionellen Intelligenztests liegt die Annahme zugrunde, dass alle Personen dieselben Lernchancen haben und dieselben Lernerfahrungen gemacht haben. Davon ausgehend würden Unterschiede in Intelligenztestscores auch Unterschiede in der Fähigkeit, von Lernerfahrungen zu profitieren, reflektieren. Somit würden Intelligenztestscores zwischen Personen mit hoher und geringer Lernfähigkeit differenzieren. Allerdings ist bekannt, dass in verschiedenen Populationen (z.B. in Bezug auf die kulturelle oder die sozioökonomische Herkunft sowie auf Lernschwierigkeiten oder geistige Behinderung) unterschiedliche Lernerfahrungen existieren. Insbesondere bei Personen mit geistiger Behinderung werden die Bildungsziele oftmals auf soziale und praktische Fähigkeiten begrenzt, das Unterrichten von schulleistungsbezogenen Fähigkeiten und Wissen wird oftmals vernachlässigt (Hessels & Hessels-Schlatter, 2010; Hessels-Schlatter, 2002). Darüber hinaus liefern Intelligenztestergebnisse zwar Informationen über die Ergebnisse des bislang Gelernten, aber keinerlei Informationen über aktuelles Lernen oder gar über die darin einbezogenen kognitiven Prozesse. Folglich ist die Vorhersagevalidität dieser Tests für Personen mit Lernschwierigkeiten,

(leichter) geistiger Behinderung oder anderen Risikogruppen, stark eingeschränkt. Zudem eignen sich diese Tests nicht für die Evaluierung von kognitiven Fähigkeiten (Hessels, 1997; Hessels-Schlatter, 2002; Resing, 2006; Tiekstra et al., 2009).

Im Gegensatz dazu zeigen dynamische Tests zur Erfassung von Lernfähigkeiten (LF) zuverlässige und valide Messungen der allgemeinen Intelligenz von Personen mit Lernschwierigkeiten oder (leichter) geistiger Behinderung und erwiesen sich als gute Prädiktoren zukünftiger Lernerfolge (Hessels & Hessels-Schlatter, 2010; Lidz & Elliott, 2000; Van der Aalsvoort, Resing & Ruijsenaars, 2002). Diese Testverfahren, sogenannte Lernpotentialtests oder Lernfähigkeitstests (LT), beruhen auf unterschiedlichen Durchführungsweisen, aber alle beinhalten eine Lernphase als fixen Bestandteil des Verfahrens. Insofern erfasst ein LT nicht nur das, was Personen bislang gelernt haben, sondern auch, wie leicht neues Wissen erworben werden kann. Bei der Vorhersage von schulbezogenem Lernerfolg zeigte die Forschung, dass LTs im Vergleich zu traditionellen IQ-Tests denselben Anteil an Varianz aufklären, LTs klären darüber hinaus aber zusätzlich einen substanziellen Anteil an Varianz auf (z.B. Beckmann, 2001; Hessels & Hessels-Schlatter, 2002; Tiekstra et al., 2009).

Hessels Analogical Reasoning Test (HART)

Der Hessels Analogical Reasoning Test (HART; siehe dazu Hessels, 2009; Hessels, Berger & Bosson, 2008; Tiekstra et al., 2009) ist ein Test zur Erfassung der LF und beinhaltet eine Trainingsphase, auf welche die Testphase folgt. Sein Nutzen für die Klassifizierung und Vorhersage von Fähigkeiten wurde bereits in unterschiedlichen Populationen, wie beispielsweise Kinder mit Migrationshintergrund, Schulkinder mit Lernschwierigkeiten und Jugendliche mit geistiger Behinderung, gezeigt. Die Aufgaben zum schlussfolgernden Denken werden

dabei ähnlich wie in Intelligenztests in Form von Matrizen präsentiert. Aufgabe der Probandinnen und Probanden ist es, die Regel(n) zu erkennen, welche die dargebotenen Matrizenelemente verbinden. Während oben die Aufgabe präsentiert wird, gilt es unten, die richtige Lösung aus vier möglichen Alternativen zu identifizieren. Abbildung 1 zeigt ein Beispiel der Matrizenaufgabe. In der Trainingsphase werden Prozesse und Regeln geübt, welche zum Lösen von Aufgaben dieser Art benötigt werden. Beispielsweise wird die systematische Analyse der Matrize und der Lösungsalternativen geübt und es werden Vergleiche einzelner Elemente innerhalb der Matrize durchgeführt. Zudem werden Regeln gesucht, welche für die Ableitung der Beziehungen der einzelnen Matrizenelemente und der Lösungsalternativen benötigt werden. Diese Trainingsphase erwies sich als nötig, da jüngere Kinder, aber auch Jugendliche mit Lernschwierigkeiten oder geistiger Behinderung oftmals nicht verstehen, was von ihnen bei der Aufgabenbearbeitung erwartet wird (siehe z.B. Hessels & Hessels-Schlatter, 2008). Außerdem neigen Personen mit Lernschwierigkeiten und/oder geistiger Behinderung dazu, kognitive Aufgaben sehr passiv und unvollständig zu erkunden sowie die Informationen nur oberflächlich und mangelhaft zu enkodieren (Ellis & Dulaney, 1991; McConaghy & Kirby, 1987; Paour, 1992; Wong, 1996). Folglich wenden Risikokinder beim Bearbeiten der Aufgaben oftmals nicht schlussfolgerndes Denken an. Wishart und Duffy (1990) konnten beispielsweise zeigen, dass die Anwendung der Bayley Scales of Infant Development (BSID; Bayley, 1969) bei Kindern mit Down Syndrom zu unreliablen Ergebnissen führt. Wenngleich die Gesamtergebnisse von zwei Durchgängen relativ ähnlich waren, so konnten viele Kinder in der zweiten Testphase Items nicht mehr richtig lösen, welche sie jedoch in der ersten Testphase korrekt gelöst hatten (und umgekehrt). Auch Hessels (2009) berichtete in den Pretestergebnissen des HARTs von geringer Über-

einstimmung in den korrekten Antworten von Kindern im Alter von fünf bis sieben Jahren sowie bei Schulkindern, welche eine Sonderschule besuchten. Beispielsweise lösten Kinder sehr schwierige Items, konnten jedoch ganz einfache Items nicht lösen. Ähnliche Ergebnisse wurden von dem Autor bei Personen zwischen 16 und 54 Jahren mit moderater bis schwerer geistiger Behinderung beim Lösen der Matrizen des CPMs (Raven, 1965a) gefunden. Das durchschnittliche mentale Alter der Personen lag bei 6 Jahren und 7 Monaten. Wenngleich die Reliabilität und die Stabilität zufriedenstellende Werte indizierte, so waren etwa 20% der gelösten Items nicht stabil, das bedeutet, dass die Items entweder beim Prä- oder Posttest gelöst wurden, jedoch wurden sie nicht konstant in beiden Durchgängen gelöst (Hessels, Rojas & Beaufort, 2012). Dies weist eher darauf hin, dass die Probandinnen und Probanden die Antworten geraten haben, anstelle schlussfolgerndes Denken anzuwenden. Somit sind sowohl die Testreliabilität als auch die Konstruktvalidität des Tests nicht gewährleistet (Beckmann, 2001, 2006; Hessels, 2009; Hessels-Schlatter, 2002).

Bei der Anwendung von LTs wird angenommen, dass Kinder in der Trainingsphase lernen, die gewünschten Prozesse und Regeln im schlussfolgernden Problemlösen zu benutzen und dass die inter-individuelle Variabilität in der Durchführung dieser Prozesse und Anwendung der Regeln beim Posttest die Lernfähigkeit anzeigt. Die einzigen Indikatoren, welche jedoch verwendet werden, um Änderungen anzuzeigen, sind die Differenzen zwischen den Prä- und Posttestscores und die relativ geringe Korrelation zwischen Prä- und Posttest. Dabei werden Kontrollgruppen einbezogen, um Trainingseffekte kontrollieren zu können. Die relativ geringe Korrelation zwischen Prä- und Posttest zeigt an, dass Probandinnen und Probanden in unterschiedlicher Weise vom Training profitierten. Manche erzielten einen hohen Zuwachs und zeigten somit eine hohe Lernkapazität, andere hin-

gegen nur einen kleinen oder überhaupt keinen Zuwachs und zeigten demzufolge auch eine geringe Lernkapazität. Je größer diese Änderung ist, desto niedriger ist die Korrelation zwischen Prä- und Posttest. Die hier genannten Indikatoren sind jedoch nur implizite Hinweise der Veränderung. Eine Möglichkeit, um zu prüfen, ob im Rahmen der Trainingsphase wirklich Veränderungen im Lösungssuchverhalten stattgefunden haben, ist Blickbewegungen aufzuzeichnen und diese mit der Leistung im schlussfolgernden Denken in Verbindung zu setzen.

Schlussfolgerndes Denken und Blickbewegungsanalysen

Zahlreiche Studien befassen sich mit dem Zusammenhang zwischen Lesen und visueller Wahrnehmung (z.B. Zhao et al., 2014). Aktuell werden insbesondere Probandinnen und Probanden mit Autismus ins Blickfeld genommen (z.B. Guillona, Hadjikhani, Badaula & Rogé, 2014). Hinsichtlich des Zusammenhangs zwischen Blickbewegungen und dem Bearbeiten von Aufgaben zum schlussfolgernden Denken gibt es jedoch nur wenig Befunde.

Bethell-Fox, Lohman und Snow (1984) demonstrierten mittels der Analyse von Blickbewegungen, dass Schülerinnen und Schüler in Abhängigkeit der Itemschwierigkeiten und ihrer Fähigkeit unterschiedliche Problemlösungsstrategien benutzten. Personen mit geringeren Fähigkeiten wendeten dabei häufiger Ausschussverfahren von Lösungsalternativen an, indem Elemente der Matrize mit den Alternativen verglichen wurden. Im Vergleich dazu zeigten Personen mit höherer Fähigkeit häufiger elaborierte Strategien wie das sog. konstruktive Matching. Dies bezeichnet das Analysieren der Information der Matrize, um eine mentale Repräsentation der benötigten Antwort zu finden, welche dann mit den Lösungsmöglichkeiten verglichen wird. Bethell-Fox und Kollegen (1984) zufolge hängt die Itemschwierigkeit mit der Anzahl der zu betrachtenden Elemente wie auch mit der An-

zahl der Antwortalternativen zusammen, was zu einer kognitiven Überbeanspruchung führen kann. Jedoch konnten die Autoren den Befund von Sternberg (1977), dass fähigere Personen sich mehr Zeit für die Enkodierung der Matrizen nehmen, nicht replizieren. Abbildung 1 zeigt ein (fiktives) Beispiel der Blickbewegungen einer Person in einer HART-Aufgabe, welche verstärkt das Ausschlussverfahren von Lösungsalternativen anwendet (oben) und einer Person, welche konstruktives Matching anwendet (unten). In der oberen Bildhälfte geht der Blick vom Startpunkt, welcher vor der Aufgabendarbietung gezeigt wurde (und

verschwindet, bevor die Aufgabe präsentiert wird), zum Element A der Matrize (Schritt 1), danach geht der Blick direkt zu den Lösungsalternativen (Schritt 2) und erst dann wird die Matrize wieder analysiert (Schritt 3). Insgesamt sieht man, dass der Blick häufig zwischen Matrize und Lösungsalternative springt, ohne, dass die Matrize zuvor systematisch analysiert wurde. Bei den Blickbewegungen in der unteren Abbildung erkennt man, dass der Blick vom Startpunkt in die Matrize geht und zuerst die einzelnen Matrizenelemente analysiert werden (Schritte 1-5), bevor die Lösungsalternativen angeschaut werden (ab Schritt 6). Darüber

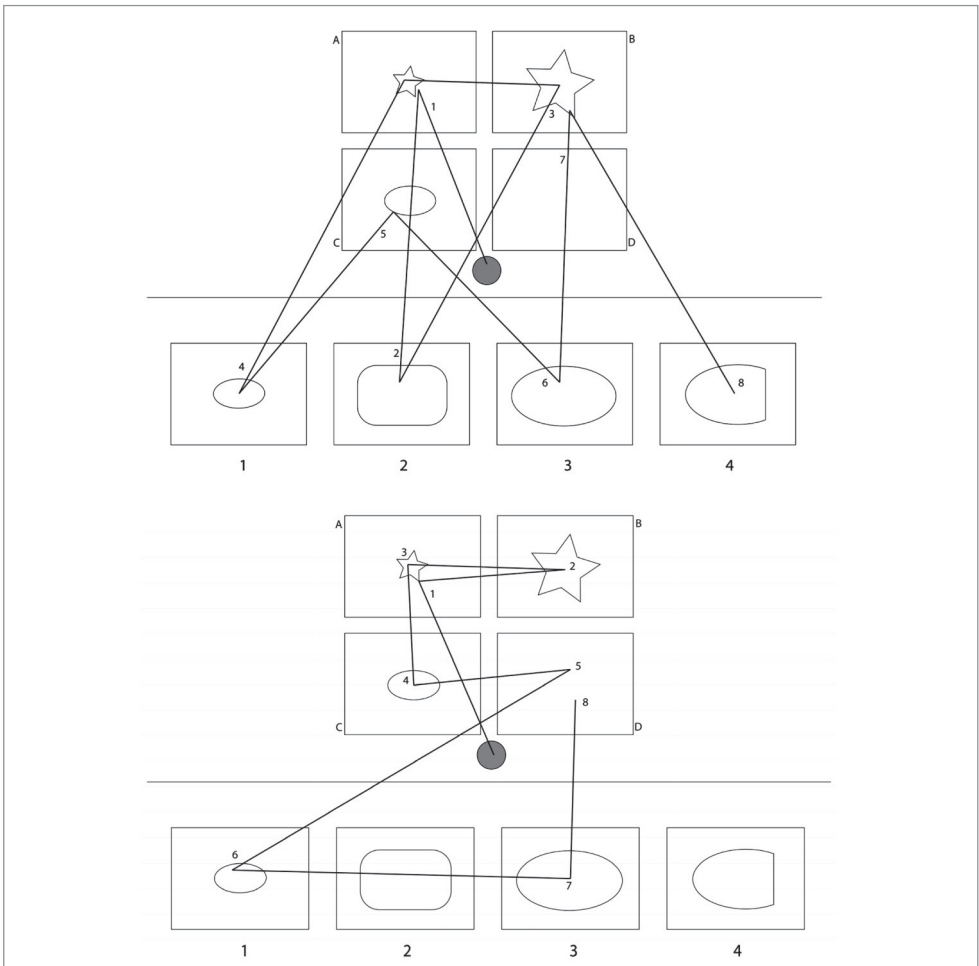


Abbildung 1. Beispiel einer Matrizenaufgabe des HARTs mit fiktiven Blickbewegungen (inkl. Startpunkt) unter der Verwendung unterschiedlicher Strategien

hinaus sieht man, dass der Proband nur noch jene Lösungsalternativen (1 und 3) betrachtet, welche er (nach Sichtung der Matrize) als relevant erachtete.

Vigneau, Caissie und Bors (2006) verwendeten 14 Items des Raven Advanced Progressive Matrices Tests (APM; Raven, 1965a) bei Universitätsstudierenden und kamen zu dem Ergebnis, dass zwei Faktoren individuelle Unterschiede im schlussfolgernden Denken erklären können: die Antwortgeschwindigkeit und die verwendeten Lösungsstrategien. Probandinnen und Probanden mit hoher Leistung verwendeten mehr Zeit, um die Matrize zu analysieren und zeigten relativ lange Fixierungen und Sakkaden, welche sich auf die Matrize bezogen, beispielsweise, um diese komplett zu enkodieren (siehe dazu auch Sternberg, 1977), bevor die erste Sakkade zu den Antwortalternativen ging. Die Ergebnisse von Vigneau et al. (2006) stehen mit jenen von Bethel-Fox et al. (1984) im Einklang. Die individuelle Art, wie Personen die Aufgabe bearbeiten, entweder durch das Ausschlussverfahren falscher Lösungsmöglichkeiten oder dem konstruktiven Matching-Vergleich, ist für den Erfolg entscheidend. Vigneau et al. (2006) zeigten, dass Probandinnen und Probanden, welche relativ viel Zeit damit verbrachten die Matrize zu untersuchen, diese Information verwendeten, um die richtige Lösungsalternative zu identifizieren. Zudem zeigte sich, dass diese Personen weniger Zeit für die Analyse von einfachen Items verwendeten, wodurch der Problemlöseprozess beschleunigt wurde.

Bethge, Carlson und Wiedl (1982) untersuchten die Bedeutung von LTs anhand der Raven's Coloured Progressive Matrizen (CPM; Raven, 1965b) bei Kindern der dritten Schulstufe. Bethge und Kollegen teilten ihre Stichprobe in drei Subgruppen. Die erste Gruppe erhielt den CPM mit den regulären Instruktionen. Die zweite erhielt ein elaboriertes Feedback (z.B. ob ihre Antwort korrekt war und zudem detaillierte Informationen darüber, warum diese korrekt oder inkorrekt war). Die dritte Gruppe wurde ge-

beten, die Matrizen zu beschreiben und zu erklären, warum sie sich für eine bestimmte Antwortalternative entschieden, bevor sie eine Lösungsmöglichkeit auswählten. Dabei stellten die Autoren fest, dass sich die Blickbewegungen veränderten. Dies bestätigt, dass sich das Problemlöseverhalten in den drei Bedingungen unterschied. Die Dauer der Fixierung, die Anzahl der Fixierungen und auch die Anzahl der Vergleiche zwischen der Matrize und den Antwortalternativen und jene zwischen den einzelnen Antwortalternativen waren in der zweiten und dritten Gruppe erhöht. Auch die Performanz erhöhte sich entsprechend. Insgesamt legen die Ergebnisse nahe, dass durch das dynamische Testen mehr geeignete und geplante Strategien induziert werden. Auch die Ergebnisse von Dillon (1985) bestätigen diese Annahme.

Blickbewegungsanalysestudien mit dem HART

Hessels, Vanderlinden und Rojas (2011) führten eine Pilotstudie mit einer computer-gestützten Version des HARTs durch, um herauszufinden, ob sich durch ein Training die Bearbeitungsschritte zum Lösen der Items bei Schülerinnen und Schülern mit und ohne Lernbehinderung ($N = 11$) verändern. Die Analyse der Blickbewegungen sollte dabei Hinweise für die Konstruktvalidität des Tests liefern. Die Ergebnisse der Pilotstudie zeigten erwartungsgemäß eine Leistungssteigerung zwischen dem Prä- und dem Posttest (Cohen's $d = 0.88$). Zusätzlich reduzierte sich jedoch die Zeit für die Lösungsfindung (Cohen's $d = -0.58$). Daraus lässt sich schlussfolgern, dass das Problemlösen der teilnehmenden Personen durch das Training effizienter wurde. Viel relevanter scheint jedoch zu sein, dass die Probandinnen und Probanden nach dem Training deutlich mehr Zeit (ca. 50% länger) mit der Analyse der Matrize (Enkodierung; siehe Abbildung 2) verbrachten, bevor die Antwortalternativen analysiert wurden (Cohen's $d = 0.49$). Die Blickbewegungsanaly-

sen ließen zudem erkennen, dass die Schülerinnen und Schüler die Details der Matrize intensiver analysierten (Cohen's $d = 0.99$) (dies lässt sich daraus schließen, dass für die erste Analyse der Matrize mehr Zeit verwendet wurde, siehe dazu Abbildung 2). Darüber hinaus wurden im Posttest häufiger „intelligenter“ Vergleiche innerhalb der Matrize vorgenommen (Cohen's $d = 0.77$), da die einzelnen Matrizenbestandteile systematischer miteinander verglichen wurden. Dies wiederum spricht für die häufigere Anwendung von konstruktiven Matchingstrategien. Darüber hinaus zeigten sich starke Zusammenhänge zwischen der Dauer der Analyse der Details der Matrizen sowie der Anzahl der intelligenten Vergleiche mit dem Testscore des Posttests, was die Konstruktvalidität des Tests bestätigt.

Schorno (2013) führte eine zweite Studie mit 44 Jugendlichen mit leichter geistiger Behinderung durch und bestätigte die Ergebnisse von Hessels und Kollegen (2011). Mittels Kontrollgruppendesign stellte sie fest, dass die Experimentalgruppe, welche zwischen dem Prä- und Posttest ein Training erhielt, sich stärker verbessert (Cohen's $d = .77$). Zudem stieg auch in dieser Studie die verwendete Zeit für die Enkodie-

rung an (Cohen's $d = .84$). Der Zusammenhang zwischen der verwendeten Zeit für die Enkodierung und dem Testergebnis stieg in der Experimentalgruppe stark an, während er in der Kontrollgruppe sank.

In einer weiteren Studie ($N = 37$) zeigte Zuretti (2014), dass die Trainingseffekte bei Kindern ohne Behinderung, welche jedoch das idente mentale Alter wie die Jugendliche mit geistiger Behinderung aufwiesen, ähnlich waren. Allerdings benötigten die Jugendlichen mit geistiger Behinderung mehr Zeit und führten auch mehr Vergleiche durch als die Kinder ohne Behinderung. Dies könnte dadurch erklärt werden, dass Personen mit geistiger Behinderung oftmals ein geringes Arbeitsgedächtnis besitzen und demzufolge häufiger Vergleiche wiederholen müssen, um die Information im Arbeitsgedächtnis präsent zu haben. Die unterschiedlichen Bearbeitungszeiten zwischen der Experimental- und Kontrollgruppe waren jedoch nicht signifikant. Die Enkodierung hing nur in der Experimentalgruppe signifikant mit den Ergebnissen im HART zusammen, sowohl im Prä- als auch im Posttest. Insgesamt sind die Ergebnisse von Zuretti (2014) schwierig zu interpretieren, was auf messinstrumentsspezifische Probleme

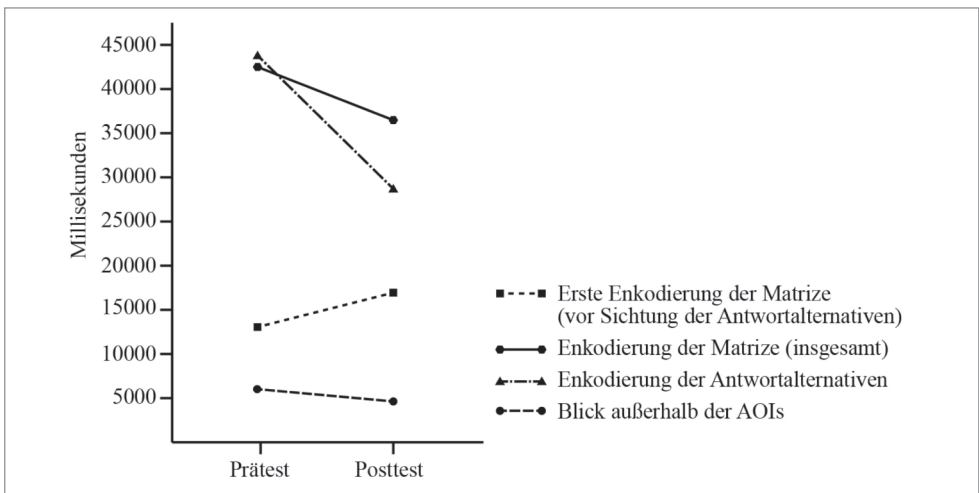


Abbildung 2: Ergebnisse der Studie von Hessels et al. (2011) in Bezug auf die Analyse der Dauer der Blicke in unterschiedlichen AOIs

me beim Eye-Tracking zurückgeführt werden kann.

Messinstrumentenspezifische Probleme beim Eye-Tracking

Die Genauigkeit der spezifischen Eye-Tracking-Systeme wird durch einen Fehlerbereich definiert, welcher bei der Bestimmung der AOIs beachtet werden muss (z.B. $0.4\text{--}0.7^\circ$ bei der Verwendung des Eye-Trackers Tobii T60). Demzufolge muss zwischen den einzelnen Elementen des AOIs auch ausreichend freier Platz sein, um die Messgenauigkeit einzubeziehen. In den zuvor genannten Studien wurde dies so umgesetzt, dass die einzelnen Elemente in klar voneinander getrennten Kästchen präsentiert wurden (siehe dazu auch Abbildung 1). Die entsprechenden Elemente des AOIs wurden dann etwas größer definiert als die Kästchengröße, um sicherzugehen, dass die Ungenauigkeit der Aufzeichnung dennoch der Analyse des richtigen Kästchens entspricht. Die Genauigkeit des Eye-Tracking-Systems erlaubt es dann (oder eben nicht), dass auch der Vergleich von Details innerhalb eines Elements analysiert werden kann. Dabei muss auch beachtet werden, dass sich die Genauigkeitsangabe von Eye-Tracking-Systemen der Hersteller zumeist auf optimale Laborbedingungen bezieht, welche in der Regel weit entfernt von den Bedingungen sind, die bei Forscherinnen und Forschern vorliegen (Hessels, Cornelissen, Kemner & Hooze, 2014). Dies führt nicht nur zu einer geringeren Genauigkeit der Messung, sondern häufig auch zum Verlust von wichtigen Daten. Dies ist beispielsweise der Fall, wenn Studien mit jüngeren Kindern oder Personen mit einer Behinderung durchgeführt werden. Jüngere Kinder sowie Personen mit Behinderung sitzen in der Regel nicht regungslos vor dem Bildschirm, sondern bewegen sich häufig. Zudem wenden sie ihren Blick aufgrund von mangelnder Konzentration außerhalb des Bildschirms. Hessels et al. (2014) evaluierten acht unterschiedliche Einstellungen von Eye-Tracking-

Systemen und zeigten dabei, dass Kopfbewegungen zu Verzerrungen in der Blickbewegungsaufzeichnung oder sogar zu einem kompletten Datenverlust führten. Zudem war es für das System nicht immer möglich, der Blickbewegung wieder zu folgen, wenn diese einmal verloren wurde. In den Studien von Hessels et al. (2011), Schorno (2013) und Zuretti (2014) führte dies beispielsweise zum Ausschluss einiger Probandinnen und Probanden mit geistiger Behinderung oder auch von jüngeren Kindern (z.B. Schorno, 2013: 13 von 57 Probandinnen und Probanden wurden aufgrund fehlender Blickbewegungsaufzeichnungen ausgeschlossen).

Des Weiteren kann es bei bestimmten Eye-Tracking-Systemen passieren, dass die Aufzeichnungsfrequenz verringert wird, sobald das System den Blick verliert. Erfolgt in einem nächsten Schritt eine automatische Bearbeitung der Daten, so ist diese stark verzerrt. Bei Hessels et al. (2014) wurde beispielsweise gezeigt, dass die Aufzeichnung von einer Frequenz von 250 Hz (250 Messungen pro Sekunde) beim Verlieren des Blicks automatisch auf etwa 20 Hz reduziert wurde. Da die Anzahl der Messungen pro Sekunde stark schwanken können, ist es auch problematisch, mit fehlenden Aufzeichnungspunkten umzugehen. In der Regel wird bei Missings nämlich davon ausgegangen, dass, wenn z.B. der Blick beim Aufzeichnungszeitpunkt 1 am Punkt x ist, und beim Zeitpunkt 4 auch auf dem Punkt x ist, dieser auch bei den Zeitpunkten 2 und 3 (wo keine Aufzeichnungen möglich waren) am Punkt x war. Die Frequenz der Aufzeichnungen muss demnach auch bei der Zeitspanne, in der die AOIs betrachtet werden, berücksichtigt werden, um valide Daten zu bekommen (Hessels et al., 2014). Die oben genannten Studien, welche sich mit jüngeren Kindern und Jugendlichen mit geistiger Behinderung befassten, waren auch damit konfrontiert, dass die Gesamtzeit bis zur Lösungsfindung nicht immer dem entsprach, was erwartet wurde. Dies könnte mit der Verringerung

der Aufzeichnungsfrequenz in Zusammenhang stehen.

Konklusion

Zusammenfassend kann festgehalten werden, dass die bisherigen Forschungsergebnisse der Eye-Tracking-Studien die Validität von Lerntests bestätigten. Allerdings muss gleichzeitig betont werden, dass die Analyse von Eye-Tracking-Ergebnissen in Bezug auf Genauigkeit und Validität einige Probleme aufweist und die Ergebnisse insofern in weiteren Studien repliziert werden sollten. Im Einklang mit den Ergebnissen von Hessels et al. (2014) wird vorgeschlagen, dass die Robustheit spezifischer Eye-Tracking-Systeme analysiert werden muss, um ausreichend präzise und valide Daten zu erhalten. Im Bezug auf die bisherigen Studien muss bei der Interpretation von Forschungsergebnissen einschränkend auf die (Un-)Genauigkeit der Messungen in Bezug auf die sogenannten AOIs und auch auf Datenverlust bei nicht optimalen Testbedingungen verwiesen werden. In Abhängigkeit an die Daten, welche über die Blickbewegungsanalysen gewonnen werden und die Art der Stichprobe, sollte man im Vorfeld die spezifischen Empfehlungen betrachten, um die Datenqualität hoch zu halten. Beispielsweise führen Sasson und Elison (2012) Empfehlungen für die Durchführung von Eye-Tracking-Analysen bei Kindern mit Autismus-Spektrum-Störung an. Zudem sollte man das jeweils situationsadäquate Eye-Tracking-System auswählen (siehe dazu Hessels et al., 2014).

Literaturverzeichnis

Bayley, N. (1969). *Bayley Scales of Infant Development*. New York: Psychological Corporation.

Beckmann, J.F. (2001). *Zur Validierung des Konstrukts des Intellektuellen Veränderungspotentials [The validation of the*

learning potential construct]. Berlin: Logos Verlag.

- Beckmann, J.F. (2006). Superiority: Always and everywhere? On some misconceptions in the validation of dynamic testing. *Educational and Child Psychology, 23*, 35–49.
- Bethell-Fox, Ch., Lohman, D. & Snow, R. (1984). Adaptive reasoning: componential and eye movement analysis of geometric analogy performance. *Intelligence, 8*, 205-238.
- Bethge, H., Carlson, J.S. & Wiedl, K.H. (1982). The effects of dynamic assessment procedures on Raven Matrices performance, visual search behavior, test anxiety and test orientation. *Intelligence, 6*, 89-97.
- Budoff, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency, 72*, 404-411.
- Dillon, R.F. (1985). Eye movement analysis of information processing under different testing conditions. *Contemporary educational psychology, 10*, 387-395.
- Ellis, N.R. & Dulaney, C.L. (1991). Further evidence for cognitive inertia of persons with mental retardation. *American Journal on Mental Retardation, 95*, 613-621.
- Gaschler, R., Marewski, J.N. & Frensch, P.A. (in press). Once and for all - How people change strategy to ignore irrelevant information in visual tasks. *The Quarterly Journal of Experimental Psychology, 68*(3), 543-567.
- Guillona, Q., Hadjikhani, N., Baduela, S. & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews, 42*, 279-297.
- Hessels, M.G.P. (1997). Low IQ but high learning potential: Why Zeyneb and Moussa do not belong in special education. *Educational and Child Psychology, 14*, 121-136.
- Hessels, M.G.P. (2009). Estimation of the predictive validity of the HART by means of a dynamic test of geography. *Journal of*

- Cognitive Education and Psychology*, 8, 5-21.
- Hessels, M.G.P., Berger, J.L. & Bosson, M. (2008). Group assessment of learning potential of pupils in mainstream primary education and special educational classes. *Journal of Cognitive Education and Psychology*, 7, 43-69.
- Hessels, M.G.P. & Hessels-Schlatter, C. (2002). Learning potential in immigrant children in two countries: The Dutch and the Swiss-French version of the Learning Potential Test for Ethnic Minorities. In D.G.M. van der Aalsvoort, W.C.M. Resing & A.J.J.M. Ruijsenaars (Eds.), *Learning potential assessment and cognitive training (Vol. 7). Actual research and perspectives in theory building and methodology* (pp. 227-245). New York: Elsevier Inc.
- Hessels, M.G.P. & Hessels-Schlatter, C. (2008). Pedagogical principles favouring the development of reasoning in people with severe learning difficulties. *Educational and Child Psychology*, 25, 66-73.
- Hessels, M.G.P. & Hessels-Schlatter, C. (2010). L'apport des tests d'apprentissage dans l'évaluation des capacités intellectuelles des élèves avec difficultés scolaires. In M.G.P. Hessels & C. Hessels-Schlatter, *Evaluation et intervention auprès d'élèves en difficultés*. Bern: Peter Lang.
- Hessels, M.G.P., Rojas, H. & Beaufort, N. (2012, June). *Intra-individual variability in Raven CPM in people with intellectual disabilities*. Paper presented at the 9th European Conference on Psychological Theory and Research on Intellectual and Developmental Disabilities (ECIDD). Trieste, Italy.
- Hessels, M.G.P., Vanderlinden, K. & Rojas, H. (2011). Training effects in dynamic assessment: a pilot study of eye movement as indicator of problem solving behaviour before and after training. *Educational and Child Psychology*, 28(2), 101-113.
- Hessels, R.S., Cornelissen, T.H.W., Kemner, C. & Hooge, I.T.C. (2014). Qualitative tests of remote eyetracker recovery and performance during head rotation. *Behavior Research Methods*, DOI 10.3758/s13428-014-0507-6.
- Hessels-Schlatter, C. (2002). A dynamic test to assess learning capacity in people with severe impairments. *American Journal on Mental Retardation*, 107, 340-351.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H. & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.
- Lidz, C.S. & Elliott, J. (2000) *Dynamic assessment: Prevailing models and applications*. New York: Elsevier.
- McConaghy, J. & Kirby, N.H. (1987). Using the componential method to train mentally retarded individuals to solve analogies. *American Journal of Mental Retardation*, 92, 243-254.
- Paour, J.-L. (1992). Induction of logical structures in the mentally retarded: An assessment and intervention instrument. In Haywood, H.C., Tzuriel, D. (Eds.), *Interactive assessment*. New-York: Springer-Verlag.
- Raven. J.C. (1965a). *Advanced Progressive Matrices*. London: Lewis.
- Raven. J.C. (1965b). *Coloured Progressive Matrices*. London: Lewis.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 3, 372-422.
- Resing, W.C.M. (2006). Using children's 'ability to learn' in diagnosis and assessment. *Educational and Child Psychology*, 23, 6-11.
- Sasson, N. J. & Elison, J. T. (2012). Eye Tracking Young Children with Autism. *Journal of Visualized Experiments*, 61, 3675. Verfügbar unter: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3460581/> [20.03.2015]
- Schorno, S. (2013). *Analyse de l'effet de l'entraînement du raisonnement analogique: utilisation du capteur du mouvement oculaire dans une procédure d'évaluation dynamique*. Unveröffentlichte Masterarbeit an der Universität Genf. Verfügbar

- unter: <http://archive-ouverte.unige.ch/unige:30693> [22.3.2015].
- Sternberg, R.J. (1977). Component processes in analogical reasoning. *Psychological Review*, 84, 353-378. Tiekstra, M., Hessels, M.G.P. & Minnaert, A.E.M.G. (2009). Learning capacity in adolescents with mild intellectual disabilities. *Psychological Reports*, 105, 1-12.
- Van der Aalsvoort, G.M., Resing, W.C.M. & Ruijssenaars, A.J.J.M. (Eds.). (2002). *Learning potential assessment and cognitive training. Actual research and perspectives in theory building and methodology*. New York: Elsevier.
- Vigneau, F., Caissie, A.F. & Bors, D.A. (2006). Eye movement analysis demonstrated strategic influences on intelligence. *Intelligence*, 34, 261-272.
- Wishart, J.G. & Duffy, L. (1990). Instability of performance on cognitive tests in infants and young children with Down's syndrome. *British Journal of Educational Psychology*, 59, 10-22.
- Wong, B.Y.L. (1996). *The ABCs of learning disabilities*. New York, NY: Academic Press.
- Zhao, F., Schnotz, W., Wagner, I. & Gaschler, R. (2014). Eye tracking indicators of reading approaches in text-picture comprehension. *Frontline Learning Research*, 2(4), 46-66.
- Zuretti, C. (2014). *Le test d'apprentissage HART: l'effet de l'évaluation dynamique sur des personnes avec et sans déficience intellectuelle*. Unveröffentlichte Masterarbeit an der Universität Genf. Verfügbar unter: <http://archive-ouverte.unige.ch/unige:44145> [22.3.2015].
- Marco G.P. Hessels**
Universität Genf
Fakultät der Psychologie und
Erziehungswissenschaften
Abteilung Sonder- und Heilpädagogik
Boulevard du Pont d'Arve, 40
1205 Genf
Schweiz
Marco.Hessels@unige.ch
- Susanne Schwab**
Universität Bielefeld
Fakultät für Erziehungswissenschaft
AG Schulpädagogik und Allgemeine
Didaktik
Universitätsstraße 25, Q1-108
33615 Bielefeld