

**Rezension für Tests und Arbeitsmittel
für den Anwendungsbereich der Berufs-, Studien- und Laufbahnberatung in der Schweiz**

Verfahren

Raven's Progressive Matrices 2, Clinical Edition (Raven's 2)

Raven, J., Rust, J., Chan, F., & Zhou, X. (2018). Raven's 2 Progressive Matrices, Clinical Edition (Raven's 2). Pearson. Deutsche Überarbeitung durch Pearson (2019).

Kategorie
Leistungstest

1 Beschreibung

1.1 Fragestellung / Anwendungsbereich

Das eindimensionale Intelligenztestverfahren dient sowohl im unteren als auch im oberen Leistungsbereich zur nonverbalen Beurteilung von kognitiven Fähigkeiten. Bei den Raven's 2 handelt es sich um eine Revision der Coloured Progressive Matrices (CPM), Standard Progressive Matrices (SPM) und Advanced Progressive Matrices (APM), welche die drei Tests zu einem Verfahren vereint.

1.2 Zielgruppe (Testpersonen)

Die Raven's 2 sind zum Einsatz im Altersbereich von 4;0 bis 69;11 Jahren geeignet.

1.3 Konzept / Theoretische Grundlagen

Die Raven's Progressive Matrices (RPM) werden seit über 70 Jahren als nonverbales Mass von kognitiven Fähigkeiten in verschiedenen Bereichen angewendet. Durch reduzierte verbale Instruktionen und nonverbale Testinhalte, welche auf universellen geometrischen Formen beruhen, wird der Einfluss unterschiedlicher kultureller Hintergründe auf die Testleistung minimiert. Die RPM wurden entwickelt, um eine Komponente von g, die Spearman als "eductive ability" bezeichnete, möglichst einfach und eindeutig zu erfassen. Klar denken und komplexe Probleme lösen zu können, gehören zu dieser Fähigkeit. Die "eductive ability" ist vergleichbar mit Cattells fluider Intelligenz.

Die weit verbreitete Adaption und Nutzung der RPM führten zu einer übermässigen Exposition der Items. Eines der Hauptanliegen während des Revisionsvorhabens war somit, alle Items für die Raven's 2 neu zu entwickeln, dabei aber die zugrundeliegenden kognitiven Konstrukte der RPM beizubehalten. Insgesamt wurde eine Itembank mit 329 Items erzeugt, die alle durch die CPM, SPM und APM bisherig erfassten Leistungsbereiche abdecken.

1.4 Material

Papierversion: Der komplette Test besteht aus einem Manual, 50 Antwortbogen, einem Stimulusbuch und einer Auswertungsschablone. Es wird zusätzlich eine Stoppuhr benötigt. Die Papierversion wird von Pearson herausgegeben, kann aber auch über Hogrefe bezogen werden.

Digitale Version: Die digitale Durchführung findet über die webbasierte Plattform Q-global von Pearson statt. Der Zugang ist kostenlos, die Testdurchführungen müssen jedoch kostenpflichtig erworben werden. Sobald Testdurchführungen bestellt werden, wird auf der Plattform eine

Materialdatenbank angelegt, auf der das Manual sowie Anweisungen zur digitalen Durchführung und allgemeine Informationen zu den Raven's 2 zu finden sind. Auch bei der digitalen Version wird eine Stoppuhr benötigt.

1.5 Anwendungen

1.5.1 Durchführung und Auswertung

Durchführung

Die Raven's 2 liegen in drei Formen vor: Papierform, digitale Langform und digitale Kurzform. Der altersabhängige Testeinstieg und eine Zeitbegrenzung sind vorgegeben.

Testform	Altersgruppe	Anzahl Items (Set)	Zeitbegrenzung
Papierform	4;0-8;11 / 9;0-69;11	36 (A,B,C) / 48 (B,C,D,E)	30 Min. / 45 Min.
Digitale Langform	4;0-8;11 / 9;0-69;11	36 / 48	30 Min. / 45 Min.
Digitale Kurzform	4;0-69;11	24	30 Min.

Die Raven's 2 können als Einzel- oder Gruppentest durchgeführt werden. Eine kurze Pause ist zulässig, und wird nicht zur Bearbeitungszeit gerechnet. Der Test muss immer durch die Testleitung beendet werden, auch in der digitalen Version. Die Anwendung des Verfahrens erfordert wenig verbale Instruktion und es werden keine mündlichen oder schriftlichen Antworten vorausgesetzt.

Papierversion: Die Papierform enthält unveränderte Items und der Einstieg wird anhand des Alters der Testperson festgelegt. Es ist aber zulässig, bei Personen mit einer niedrigen Leistungsfähigkeit beim Set A zu starten, auch wenn sie ausserhalb dieser Altersgruppe sind (4;0-8;11 Jahre).

Digitale Version: Bei der digitalen Durchführung wird jeder Testperson ein einzigartiges Set von zufällig ausgewählten Items vorgegeben. Ein Übungslink für Testleiter hilft, sich mit der Testdurchführung vertraut zu machen.

Auswertung

Die Auswertung der Papierform erfolgt mittels einer Schablone. Die aufsummierten Rohwerte bilden den Ausgangspunkt für die Leistungswerte, welche in Normwerte umgewandelt werden. Alternativ kann die Papierform kostenpflichtig auf Q-global ausgewertet werden. Die digitale Version wird bei der Nutzung der Plattform Q-global automatisch ausgewertet, wobei auch die Normen automatisch entsprechend der Altersgruppe gewählt werden.

1.5.2 Interpretation

Die Interpretation erfolgt anhand des IQ-Wertes und Prozentranges. Es wird empfohlen bei den IQ-Werten das Konfidenzintervall anzugeben. Es werden auch qualitative Beschreibungen der IQ-Werte angegeben. Im Manual werden drei Fallbeispiele aufgeführt.

Zusätzlich stehen Testalteräquivalente zur Verfügung, welche das Alter angeben, in dem ein entsprechender Leistungswert der mittleren Leistungsfähigkeit entspricht. Diese sollen nur mit Vorsicht interpretiert werden, da sie u.a. psychometrischen Einschränkungen unterliegen.

1.6 Gütekriterien, Literatur

1.6.1 Vom Autor geltend gemachte Gütekriterien

Objektivität:

Aufgrund der standardisierten Testdurchführung, Auswertung und Interpretation kann die Objektivität als gegeben betrachtet werden.

Reliabilität:

Die Reliabilitäten für die Papierform wurden anhand der europäischen Normstichprobe mit der Split-Half-Methode berechnet. Die Werte liegen bei der Altersgruppe von 4-16 Jahre zwischen .75 und .92 und bei der Altersgruppe von 17-69 Jahre zwischen .85 und .93. Die Reliabilitäten der digitalen Testformen gehen auf die US-amerikanische Normierungsstichprobe zurück und wurden mittels IRT-basierten Analysen berechnet. Bei der digitalen Langform liegen die Werte zwischen .85 und .90 und bei der digitalen Kurzform zwischen .79 und .81.

Die Koeffizienten für spezielle Personengruppen (Hochbegabung und intellektuelle Beeinträchtigung) liegen ungefähr in demselben Bereich wie bei der Normalstichprobe – sowohl bei der Papierform als auch bei den digitalen Testformen. Bei durchgeführten Retest-Studien liegen die korrigierten Korrelationskoeffizienten mit Ausnahme eines Wertes (.78) alle über .80.

Validität:

Es wurden Zusammenhänge der amerikanischen Version der Raven's 2 (Papierform und digitale Testformen) mit ihren Vorgängerversionen überprüft. Die IQ-Werte der Raven's 2 korrelieren hoch mit der CPM (korrigierter Wert zwischen .74 und .78), SPM (korrigierter Wert zwischen .76 und .78) und APM (korrigierter Wert zwischen .71 und .74), was darauf hinweist, dass die Verfahren dasselbe messen. Mittelwertsunterschiede fallen vor allem in Bezug auf den Flynn-Effekt erwartungsgemäss aus. Für weitere Studien in Bezug auf die konvergente und diskriminante Validität wird auf das Manual der amerikanischen Version verwiesen.

Zusätzlich wurden Validierungsstudien bei speziellen Personengruppen (Hochbegabung und intellektuelle Beeinträchtigung) durchgeführt. Als Kriterium wurden bereits vorliegende IQ-Werte verwendet. Die entsprechenden Personengruppen erreichten signifikant höhere bzw. tiefere Werte als eine Kontrollgruppe (mit sehr grossen Effektstärken).

In Bezug auf die Inhaltsvalidität wurden Itemüberprüfungen durch Experten innerhalb von Pearson sowie durch externe Experten durchgeführt.

1.6.2 *Zusatzinformationen und Beurteilung in der Literatur*

Obwohl im Manual auf das gemessene Intelligenzkonstrukt und entsprechende Missverständnisse eingegangen wird, wird bei der digitalen Auswertung das Ergebnis der Raven's 2 als "allgemeine kognitive Fähigkeit" dargestellt. Testergebnisse der Raven's 2 sollten jedoch nicht als Allgemeinintelligenz oder allgemeine kognitive Fähigkeiten beschrieben werden, da dies den Eindruck erwecken könnte, es hätte eine umfassende Abklärung der Intelligenz stattgefunden (vgl. Cordero Donoso & Renner, 2020).

Cordero Donoso und Renner (2020) finden bei eigenen Nachberechnungen einen signifikanten Effekt bei der Datenherkunft, jedoch bei kleiner Effektstärke. Ausserdem weisen sie auf vereinzelt deutliche Unterschiede in den Standardwerten hin, wenn sich der Rohwert um einen Punkt verändert. Dies betrifft vor allem den oberen (IQ > 140) und unteren (IQ < 80) Normwertbereich.

1.6.3 *Normen*

Für die Entwicklung einer europäischen Norm wurde von Mai 2018 – Juli 2019 eine Stichprobe von $N = 1200$ Personen aus sechs verschiedenen Regionen in Europa erhoben (Deutschland, Frankreich, Spanien, Grossbritannien, Niederlande/Flandern und Skandinavien). Aus jeder dieser sechs Regionen floss eine Teilstichprobe von $n = 200$ ein. Jede Stichprobe wurde anhand ausgewählter demografischer Merkmale stratifiziert, um die Repräsentativität der Teilstichprobe für diese Region zu gewährleisten. Verschiedene Untersuchungen zeigten keinen signifikanten Unterschied zwischen den Mittelwerten der unterschiedlichen Regionen.

Zur Einordnung der Testergebnisse liegen alterskorrigierte Normwerte in jeweils 21 Altersgruppen vor. Die Stichprobengrösse pro Altersgruppe umfasst 53 – 67 Personen. Gemäss dem Prinzip des "Inferential Norming" (vgl. Zhu & Chen, 2011) reicht dies für eine angemessene Qualität aus.

1.7 **Literaturhinweise**

Cordero Donoso, P. & Renner, G. (2020). *Testinformation zu den Raven's 2. Deutsche Fassung der Raven's 2 Progressive Matrices – Clinical Edition*. Ludwigsburg: Pädagogische Hochschule Ludwigsburg.

McLeod, J. W.H. & McCrimmon, A. W. (2021). Test Review. *Journal of Psychoeducational Assessment*, Vol. 39(3), 388–392. DOI: 10.1177/0734282920958220.

Raven, J., Rust, J., Chan, F., & Zhou, X. (2018). *Raven's 2 Progressive Matrices, Clinical Edition (Raven's 2)*. Pearson.

Zhu, J. & Chen, H.-Y. (2011). Utility of Inferential Norming With Smaller Sample Sizes. *Journal of Psychoeducational Assessment*, Vol. 29(6), 570-580. DOI: 10.1177/0734282910396323.

2 Einschätzung der Fachgruppe Beratung und Diagnostik

2.1 Anwendungsqualitäten

Dadurch dass die CPM, SPM und APM zu einem Testverfahren vereint wurden, messen die Raven's 2 im unteren und oberen Leistungsbereich, was sie von anderen Matrizentests unterscheidet. Das Verfahren existiert als Papierform (Langversion) oder als digitale Form (Kurz- und Langversion) und kann als Einzel- oder Gruppentest durchgeführt werden.

Bei den Raven's 2 handelt es sich um ein objektives und reliables Verfahren, welches im Bereich der Validität noch mehr Untersuchungsbedarf hat, z.B. zur konvergenten Validität mit gebräuchlichen intelligenzdiagnostischen Verfahren. Bei der aktuellen europäischen Normstichprobe wurde die Schweiz nicht berücksichtigt. Die sechs europäischen Regionen in der Stichprobe unterscheiden sich jedoch nur wenig. Die Normierungen der Vorgängerversionen sind veraltet und es wird empfohlen, diese nicht mehr zu verwenden.

Bei der Entscheidung für den Einsatz der Raven's 2 sollte miteinbezogen werden, dass das eindimensionale Verfahren nur einen Teil der Intelligenz erfasst. Wird dies berücksichtigt, stellen die Raven's 2 ein ökonomisches Verfahren zur Messung der kognitiven Fähigkeiten dar.

2.2 Gesamtbewertung für den Anwendungsbereich Berufs-, Studien- und Laufbahnberatung

Aufgrund der oben beschriebenen Einschränkungen sollten die Raven's 2 lediglich ergänzend zu anderen kognitiven Tests eingesetzt werden oder um bereits vorliegende Befunde abzusichern. Auch der Einsatz als Screening-Instrument ist denkbar. Durch die Integration von Farben wirkt der Test moderner als die SPM und APM, welche bisher in der Berufs- Studien- und Laufbahnberatung eingesetzt werden. Durch den nonverbalen Testinhalt können auch Personen mit wenig Deutschkenntnissen getestet werden. Studien zur Kulturunabhängigkeit existieren jedoch nicht. Durch die Neukonstruktion der Items wird das Problem angegangen, dass viele Items der alten Versionen bekannt sind. Ausserdem kann der Test online aufgrund des grossen Itempools mehrmals bei der gleichen Person durchgeführt werden. In der Papierversion existiert – im Gegensatz zur digitalen Version – kein Abbruchkriterium. Dies sollte berücksichtigt werden, falls der Test über die Zeitbegrenzung hinaus bearbeitet wird.

Die Anschaffungskosten der Papierversion sind relativ hoch, jedoch muss berücksichtigt werden, dass der Test drei Verfahren vereint. Die Papierversion kann online ausgewertet werden, was jedoch zusätzliche Kosten verursacht.

2.3 Anmerkungen

Die wichtigsten Änderungen gegenüber den Vorgängerversionen:

- CPM, SPM und APM wurden in einem Testverfahren integriert
- Neuer Aufgabentyp (1 x 6 Reihe) und unterschiedliche Position des fehlenden Elements
- Möglichkeit einer digitalen Anwendung (inkl. Kurztest)
- Einführung einer Abbruchregel bei der digitalen Testung
- Vereinfachte und verkürzte Testinstruktionen
- Alle Items sind mehrfarbig gestaltet

Mögliche Auswirkungen von Störungen der Farbwahrnehmung wurde bei der Itemgestaltung berücksichtigt. Die Testfairness ist lediglich bei Vorliegen einer sehr seltenen Störung der Farbwahrnehmung (Achromatopsie) beeinträchtigt.

Die Fachgruppe Beratung und Diagnostik des SDBB hat u.a. die Aufgabe, Tests und Arbeitsmittel zuhanden der Berufs-, Studien- und Laufbahnberatung zu besprechen und zu bewerten. Sie tut dies in Form dieser Rezension, die den Fachleuten aus der Berufs-, Studien- und Laufbahnberatung Entscheidungshilfen für den Einsatz von Tests und Arbeitsmitteln bieten soll.