

UDC 159.9.01
DOI: 10.34671/SCH.BSR.2021.0502.0005



**ИЗМЕНЕНИЕ И СТАБИЛЬНОСТЬ РЕЗУЛЬТАТОВ ТЕСТА «ПРОГРЕССИВНЫЕ
МАТРИЦЫ РАВЕНА» ВО ВРЕМЕНИ И МЕЖДУ КУЛЬТУРАМИ:
ИСТОРИЯ НА РУБЕЖЕ ВЕКОВ**

© Автор(ы) 2021
ORCID: 0000-0002-6185-8153
SPIN-код: 6174-3930
AuthorID: 947445

РАВЕН Джон, почетный доктор и заслуженный профессор Печского университета (Венгрия),
почетный доктор Католического Люблинского университета (Польша)
Совет по образованию Шотландии

(Шотландия, Эдинбург EH3 6QH, Грейт Кинг-стрит, jraven@ravenfamily.co.uk)

Перевод на русский язык – О.Н. Ярыгин

Аннотация. Данная статья основана на материалах, опубликованных в Cognitive Psychology (Raven, 2000b), которые ранее вошли в Руководства по прогрессивным матрицам Равена (ПМР, Raven's Progressive Matrices, RPM). Автор предлагает более развернутый обзор исследований, касающихся стабильности и изменения показателей ПМР с течением времени и развитием культуры (и их причин). Учитывая, что тесты ПМР используются уже 70 лет, выделить доказательства, касающиеся вопроса о том, насколько похожи нормы для разных культурных групп и как они менялись с течением времени не так просто, как можно было бы ожидать. Одна из причин, почему задача настолько трудна, заключается в том, что большинство исследований в этой области не только низкого качества, но и проводятся для других целей. Таким образом, большая часть исследований, в которых использовалась ПМР, была направлена на то, чтобы связать оценки ПМР с некоторыми другими переменными (такими как образовательная или профессиональная успеваемость), а не на сбор базовых нормативных данных. И как в этих исследованиях, так и в тех, которые пытались предоставить какие-то справочные данные, заинтересованные исследователи были относительно безразличны к важности выборки, несмотря на то, что это имело серьезные последствия для валидности применяемых критериев значимости. Многие исследователи пытались объяснить очевидные культурные различия, не установив сначала, насколько велики или распространены эти различия.

Ключевые слова: тест «Прогрессивные матрицы Равена», тест «Словарная шкала Милл-Хилл», стандартизация тестов, географическое и культурное влияние, изменения результатов тестов во времени, эффект Флинна, эдуктивная способность, репродуктивная способность, воспитание, образование.

**CHANGE AND STABILITY IN RAVEN'S PROGRESSIVE MATRICES SCORES OVER CULTURE
AND TIME: THE STORY AT THE TURN OF THE CENTURY**

© The Author(s) 2021

RAVEN John, Honorary Doctor and Distinguished Professor of Psychology (University of Pecs, Hungary),
Honorary Doctor (Catholic University of Lublin, Poland)
Scottish Education Council

(Scotland, Edinburgh EH3 6QH, Great King Street, jraven@ravenfamily.co.uk)

Translated into Russian – O.N. Yarygin

Abstract. This report is based upon a paper published in Cognitive Psychology (Raven, 2000b), which was itself based on material previously published in the Manuals for the Raven Progressive Matrices (RPM) – see various reference entries for Raven, Raven, and Court. It offers a more extended summary of research relating to the stability and change in RPM scores over culture and time (and their causes) than was included in the General Introduction to this book. Given that the tests have been in use for 70 years, distilling off the evidence bearing on the question of how similar are the norms for different cultural groups and how have they changed over time is not so easy as might be expected. The reasons why the data are not better than they are merit review because of their implications for future work in the area. One reason why the task is so difficult is that, as Dahlstrom (1993) noted in an article appropriately titled “Tests: Small samples, large consequences” that most of the studies in the area are not only of poor quality but also conducted for other purposes. Thus most of the research in which the RPM have been used have sought to relate RPM scores to some other variable (such as educational or occupational performance) rather than to assemble basic normative data. And, both in these studies and in those which attempted to provide some kind of reference data, the researchers concerned have been relatively indifferent to the importance of sampling ... even though this has major implications for the validity of the significance tests they sought to apply. Thus many researchers have tried to explain apparent cultural differences without first establishing just how large or pervasive those differences were.

Keywords: Raven's progressive matrices, Mill-Hill vocabulary test, standardization of tests, geographical and cultural influence, changes in results, Flynn effect, educational ability, reproductive ability, upbringing, education.

ВВЕДЕНИЕ

Учитывая, что тесты используются уже 70 лет, выделение доказательств того, насколько похожи нормы для разных культурных групп и как они менялись с течением времени, оказывается не таким простым делом, как можно было бы ожидать.

Причины, по которым данные не лучше, чем они есть, заслуживают рассмотрения из-за их последствий для будущей работы в этой области.

Одна из причин, почему задача настолько трудна, заключается в том, что, как отметил Дальстром (Dahlstrom, 1993) [15] в статье с соответствующим названием «Тесты: небольшие выборки, большие последствия» (“Tests: Small samples, large consequences”), большинство исследований в этой области не только имеют

низкое качество, но и проводится для других целей. Таким образом, большая часть исследований, в которых использовались «Прогрессивные матрицы Равена» (ПМР), была направлена на то, чтобы связать оценки ПМР с некоторыми другими переменными (такими как образовательная или профессиональная успешность), а не на сбор базовых нормативных данных. И как в этих исследованиях, так и в тех, которые пытались предоставить какие-то справочные данные, заинтересованные исследователи были относительно безразличны к важности выборки... даже несмотря на то, что это имело серьезные последствия для валидности результатов тестов, которые они пытались применить.

Таким образом, многие исследователи пытались объяснить видимые культурные различия, не установив

сначала, насколько велики или распространены эти различия.

Возрастающее осознание важности изучения изменений результатов тестов во времени

Одна из причин, по которой существует мало адекватных данных об изменении результатов с течением времени заключается в том, что, поскольку большинство психологов даже не подозревали, что оценки будут расти с течением времени, они не только не считали необходимым собирать данные, которые можно было бы использовать для изучения этого вопроса, но просто предположили, что нормативные данные, собранные в прошлом, все еще применимы. Короче говоря, они не видели необходимости повторно стандартизировать тесты.

Другой фактор, однако, заключается в том, что имеющиеся доказательства того, что результирующие баллы росли с течением времени, были неверно истолкованы (как и аналогичные данные, полученные из многочисленных перекрестных исследований, проведенных с другими тестами) как свидетельство снижения способностей после 20 летнего возраста.

Хотя некоторые исследователи (например, Owens, 1966; Bouvier, 1969; Thorndike, 1975) [53, 7, 87] действительно заметили явное увеличение оценок по некоторым компонентам «интеллекта», общий эффект, когда оценки по этим суб-тестам сочетались с другими, вряд ли можно было назвать «драматичными».

Если бы исследователи мыслили в терминах Ч. Спирмена, то есть, осознали разделение эдуктивной (продуктивной) и репродуктивной способностей, они могли бы заметить, что по существу увеличивался только один компонент «интеллекта» (а именно, эдуктивная способность), и что рост таких оценок был действительно резким. (В этом контексте мы можем отметить, что, кроме того, что в 1984 и 1987 годах Дж. Флинн задокументировал рост более тщательно, чем предыдущие исследователи, основной его вклад на самом деле состоял в том, чтобы привлечь внимание, во-первых, к скорости изменения эдуктивной способности, и, во-вторых, решительно поставить вопрос о том, привело ли это увеличение к подлинному увеличению знаний, чтобы подчеркнуть различную скорость изменения эдуктивной и репродуктивной способностей (например, Flynn, 2000) [23].)

Помимо этих основополагающих ограничений, были также существенные методологические проблемы.

Во-первых, чтобы быть значимыми, данные должны были быть разделены по возрасту, как в таблице 8.1 (которая будет более подробно объяснена позже).

Во-вторых, бимодальное и асимметричное возрастное распределение, показанное на рисунке 8.1 (представлен из работы J. Raven, 1981) [58] в сочетании с разбросом, который менялся в зависимости от возраста (также проиллюстрирован на рисунке 8.1), означал, что обычные методы обработки данных (т. е. сокращение данных до средних значений и стандартных отклонений) не могут считаться приемлемыми. Рисунок 8.1 (представлен из работы J. Raven, 1981) [58] в сочетании с разбросом, который менялся в зависимости от возраста (также проиллюстрирован на рисунке 8.1), означал, что обычные методы обработки данных (т. е. сокращение данных до средних значений и стандартных отклонений) не могут считаться приемлемыми.

Тем не менее, несмотря на обоснованность большинства этих компонентов в объяснении того, что психологи не заметили увеличения результирующих баллов, наиболее важным, вероятно, является то, что до 1979 года даже такие данные, которые были доступны, на самом деле ни у кого не вызывали подозрений увеличения (даже на Прогрессивных матрицах). Итак, давайте теперь рассмотрим те данные, которые были доступны, чтобы пролить свет на изменения во времени и культуре для молодых людей, а затем сделаем то же самое для

взрослых.

Таблица 8.1. Стандартные Прогрессивные матрицы Равена

1979. Британские перцентильные нормы для самостоятельного или группового тестирования среди молодежи (сглаженные)

Table 8.1. Standard Progressive Matrices
1979 British Percentile Norms for the Self-Administered or Group Test Among Young People (Smoothed)

Age in years (months)																				
		6½	7	7½	8	8½	9	9½	10	10½	11	11½	12	12½	13	13½	14	14½	15	15½
		6(3)	6(9)	7(3)	7(9)	8(3)	8(9)	9(3)	9(9)	10(3)	10(9)	11(3)	11(9)	12(3)	12(9)	13(3)	13(9)	14(3)	14(9)	15(3)
Percentile		to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to	to
		6(8)	7(2)	7(8)	8(2)	8(8)	9(2)	9(8)	10(2)	10(8)	11(2)	11(8)	12(2)	12(8)	13(2)	13(8)	14(2)	14(8)	15(2)	15(8)
95		33	34	37	40	42	44	46	48	49	50	51	52	53	54	54	55	56	57	57
90		30	32	35	38	40	42	44	46	47	48	49	50	51	52	53	54	54	55	55
75		22	26	30	33	36	38	41	42	43	44	45	46	47	49	49	50	50	51	51
50		16	19	22	25	31	33	36	38	39	40	41	41	42	43	44	45	46	47	47
25		13	14	15	17	22	25	28	32	33	34	36	37	38	39	41	42	42	42	42
10		10	12	12	14	16	17	19	23	27	29	31	31	32	33	35	36	36	36	36
5		9	10	11	12	13	14	15	17	22	24	25	26	27	28	29	30	33	33	33
<i>n</i>		112	138	148	174	153	166	198	172	194	187	164	164	174	185	180	196	189	191	171

Note. Based on a nationally representative sample of British schoolchildren, excluding those attending special schools (see Raven, 1981, for details). Younger and less able children were tested individually.

Примечание. Получены на основании репрезентативной национальной выборки среди британских школьников, за исключением специальных школ (Равен, 1981, [58])

Исследования молодежи

1. Исследования изменений результатов тестов с течением времени

Тест Стандартные прогрессивные матрицы Равена (СПМ) был впервые полностью стандартизирован Дж. К. Равеном для 1407 детей в Ипсвиче (Ipswich), Англия, в 1938 г. (J. C. Raven, 1941) [70]. Следующее существенное исследование Дж. К. Равена (J. C. Raven & Walshaw, 1944) [72] было проведено не для того, чтобы установить нормы для ПМР, а для сбора эквивалентных данных для теста «Словарная шкала Милл-Хилл» (МХС). Он проводился в городке Колчестер (Colchester) недалеко от Ипсвича в 1943-44 годах.

Нормы СПМ, полученные в этом исследовании, постоянно были на два исходных балла ниже, чем нормы, полученные в Ипсвиче. В 1952 году Адамс сообщил о нормах для 11621 двенадцатилетних детей в Суррее, Англия [1]. Эти данные в пределах погрешности выборки были очень похожи на нормы, полученные Равеном в 1938 в Ипсвиче. Тадденхэм, Дэвис, Дэвисон и Шиндлер (Tuddenham, Davis, Davison, Schindler, 1958) [89] в одном из немногих исследований, в которых была принята попытка установить соответствие или несоответствие британских норм условиям в Соединенных Штатах, протестировали детей нескольких школьных классов в Калифорнии. Они пришли к выводу, что британские нормы приемлемы в США. В 1963-65 годах Г. Сканес проверил 4017 детей в возрасте от 9 до 14 лет в Сент-Джонс, Ньюфаундленд. Сходство между результатами Г. Сканеса и нормами для Ипсвича 1938 года поразительно (J. Raven, 1981) [58]. Позже, в 1967 году, в Корнер-Брук (Ньюфаундленд) Г. Сканес проверил всю популяцию (2097) детей в возрасте от 10 до 14 лет. Результаты постоянно отставали от норм для Ипсвича. В 1972 году Ю. Бирт и Ф. Гилл (Byrt, Gill, 1973) [8], работая с автором, собрали данные из национальной репрезентативной выборки из 3 464 детей начальной школы в возрасте от 5 до 11 лет в Ирландской Республике. Городские нормы, казалось, соответствовали нормам для Ипсвича 1938 года, хотя показатели для сельской местности отставали от них.

Даже в 1979 году, через 40 лет после публикации теста, мало что могло свидетельствовать о заметном росте результатов. Напротив, все указывало на их стабильность.

С 1979 года история начала меняться. В этом году Крацмайер и Хорн (Kratzmeier, Horn, 1979) [37] сообщили о нормах, полученных в крупном немецком исследовании, которые были намного выше, чем в Англии в 1938 году. Данные Г. Мельхорна (Melhorn, 1980) [43] в Восточной Германии были аналогичными. Британские нормы 1979 г., составленные (с помощью гранта Совета по исследованиям в области социальных наук и при со-

действии Государственного управления переписей и обследований населения) на основе тщательно составленной выборки, предназначенной для представления как всей Великобритании, так и социально-экономических различий в ней, оказались в целом похожими на полученные в двух немецких исследованиях (J. Raven, 1981) [58].

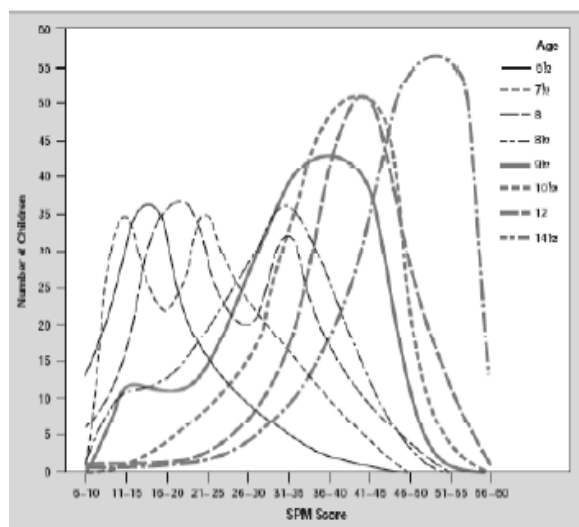


Рисунок 8.1. Стандартные прогрессивные матрицы Равена

Британская стандартизация 1979. Распределение исходных баллов по восьми возрастным группам

Холмс (1980) [31] сообщил о результатах для Британской Колумбии (Канада), которые были аналогичны национальным нормам Великобритании 1979 года, хотя и немного ниже их. Как Австралийский совет по исследованиям в области образования (см. De Lemos, 1984, 1989) [17, 18], так и Совет по исследованиям в области образования Новой Зеландии (1984) [49] сообщили о близких результатах для своих стран. Я. Ференчик (Ferjencik, 1985) [19] сообщил данные для Цветных прогрессивных матриц Равена (ЦПМР) для тогдашней Чехословакии, которые соответствовали недавно опубликованному британскому исследованию. Работа, проведенная в США Дж. Равеном (2000а) [62], показала, что, хотя общие нормы США отставали от этих новых международных показателей, нормы Уайта им соответствовали. Занг и Вонг (Zhang and Wang, 1989) [97] собрали данные для городских районов материкового Китая, которые показали, что, несмотря на предположение о таких же высоких нормах, о которых сообщал Чан (1981, 1989) [9, 10] для Гонконга, нормы для разработанной репрезентативной выборки городского населения материкового Китая близко соответствовали недавним нормам, полученным в других странах. Совсем недавно аналогичные данные были опубликованы для Польши (Jaworowska & Szustrowa, 1991; J. Raven, J.C. Raven & Court, 1998c, 2000, обновлено 2004 г.) [34, 67, 68, 69], Испании (J.C. Raven, Court, J. Raven, 1995) [71], новых школьных округов США (J. Raven, 1989) [60], Швейцарии (Martinoli, 1990; Spicher, 1993) [39, 84] и Индии (Deshpande, in J. Raven, J.C. Raven, & Court, 2000) [69].

Здесь можно отметить два замечания. Во-первых, общая о результатах британской стандартизации 1979 г., мы сами (J. Raven, 1981) [58], отмечая разницу между нормами 1938 и 1979 гг., не комментировали ее магнитуду и, не обратив внимания на то, что оценки более способных подростков приблизились к максимуму, достижимому в тесте, предположили, что увеличение произошло в основном среди менее способных. Во-вторых, учитывая сходство норм, о которых сообщали все вышеперечисленные исследователи, которые опублико-

вали данные с 1936 по 1979 год, и сходство норм 80-х годов, о которых сообщали другие авторы, чьи работы были обобщены, не было никаких намеков на то, что мы можем видеть свидетельства *непрерывного* увеличения оценок с течением времени. Просто мог быть прыжок.

2. Географические и культурные различия

Рассмотренные до сих пор исследования предполагают, что нормы для разных групп населения в определенный момент времени схожи, но в 1970-е годы каким-то образом резко выросли.

Теперь мы подведем итоги исследований, документирующих различия в нормах для молодых людей из разных географических регионов и разных культур, как в качестве самостоятельной темы, так и с целью изучения того, какой свет они могут пролить на изменения с течением времени. Вначале будут рассмотрены исследования, выявляющие большие различия между странами, а за ними последует обзор исследований различий внутри стран.

Как уже упоминалось, гонконгские нормы Чана превышают большинство уже упомянутых норм. Однако нормы, наиболее значительно превышающие их, исходят из Тайваня (Miao & Huang, 1990; Miao, 1993) [44, 45]. (Чтобы снизить вероятность неправильного толкования этих норм, следует отметить, что в ходе посещения Тайпея выяснилось, что тест ПМР не только играет важную роль в жесткой тайваньской системе оценки образования, но и все учителя получают копии теста и поощряются к «тренингу учеников с его использованием». Поэтому кажется вероятным, что «высокие» тайваньские нормы свидетельствуют не о превосходных эдуктивных способностях тайваньцев, а скорее о том, насколько мал максимальный эффект мотивации и обучения.)

С другой стороны, как также отмечалось, нормы для сельских и изолированных сообществ обычно ниже, чем для других. В данном контексте можно рассматривать вышеупомянутые нормы для Ирландской Республики и Ньюфаундленда как подтверждение этого. Другие низкие нормы для, казалось бы, хороших выборок исследуемых популяций были зарегистрированы в Бразилии (Angelini, Alves, Custodio, & Duarte, 1988) [2], Турции (Sahin & Duzen, 1994) [73], Малайзии (Chiam, 1994, 1995) [11, 12], Пуэрто-Рико (Kahn, Spears, & Rivera, 1977; J. Raven, 1989) [36, 60] и в отдаленном районе в горах Перу (см. Munoz in Raven et al., 1998b) [66].

Как подчеркивалось Дж. Равеном (1989) [60], «низкие» нормы, полученные в большинстве этих исследований, должны быть связаны с соответствующим контекстом, отмечая, что, за заметным исключением перуанских горных норм, большинство из них выше британских. 1938 г. норм. Отсюда следует, что факторы, которые были ответственны за обсуждаемое увеличение результирующих баллов с течением времени, также могли вызвать различия между культурными группами.

Более систематические исследования различий между географическими, социально-экономическими и этническими группами внутри стран были предприняты в ходе стандартизации молодежи в Великобритании и США. Поскольку и дизайн, и переменные, рассматриваемые в этих двух исследованиях, были разными, их необходимо обсуждать отдельно.

Британская стандартизация 1979 года

Британская стандартизация 1979 года проводилась в семи областях страны, которые были выбраны под руководством Государственного управления переписей и обследований населения (Government Office of Population Censuses and Surveys - OPCS) для представления всех типов территорий, в которых проводится кластерный анализ больших объемов демографических данных. Эта стандартизация показала, что социально-экономические различия внутри страны можно классифицировать (Webber, 1977) [96]. Типы районов, в которых проживает мало людей, были выбраны в избыточном объеме, чтобы иметь достаточное количество респондентов, для раз-

бienia данных по типам регионов. Позже данные были повторно масштабированы до правильных пропорций, чтобы получить общую статистику. Поэтому при анализе данных можно было использовать довольно сложные статистические процедуры. Всего было протестировано 3250 детей в возрасте от 6 до 16 лет.

Сейчас подходящий момент для объяснения формата, в котором будут отображаться данные. Многие авторы представляют свои данные с точки зрения IQ-отклонения со средним значением 100 и стандартным отклонением 15.

Этот представление, как правило, неоправданно по ряду причин, включая две, которые здесь важны: во-первых, как видно из рисунка 8.1, распределение оценок ПМР внутри возрастной группы (и, согласно личному сообщению Роберта Торндайка, подшкалы Стэнфорд-Бине), как правило, не являются гауссовыми и действительно часто являются бимодальными. Во-вторых, это не способствует исследованию того, могут ли существовать разные тенденции на разных уровнях способностей.

Чтобы избежать этих (и других) проблем, нормативные данные для ПМР всегда представлялись в виде таблиц, показывающих необработанные оценки, необходимые для достижения более, чем 5%, 10%, 25%, 50%, 75%, 90% и 95% населения определенного возраста.

Таблица 8.1, в которой представлены общие результаты британского исследования норм среди молодежи 1979 года, является типичным результатом.

Уместно еще одно пояснение. Чтобы минимизировать эффекты ошибки выборки (обычно довольно большие), цифры в этих таблицах обычно сглаживаются путем построения графика для каждой перцентильной кривой, которая, насколько это возможно, (а) выравнивает отклонения исходных (сырых) оценок, указанных выше и ниже линии, (б) придает наибольший вес наиболее надежным данным (данные для самых старших и самых молодых возрастных групп в выборке обычно искажаются такими факторами, как политика продвижения и удержания в образовании), и (с) отражает тенденции в самых надежных данных.

Необходимость сглаживания данных для минимизации эффектов случайной дисперсии и, главное, ошибок выборки можно увидеть на рисунках 8.5 и 8.9. Популяции, на которых основаны эти графики, вообще некорректно описываются как «выборки», потому что они охватывают практически полные популяции последовательных возрастных срезов бельгийских мужчин.

Тем не менее, невозможность графически отобразить данные может скрыть общие тенденции, которые так поразительны, и всегда будет соблазн сосредоточиться на объяснении колебаний вокруг общей тенденции, которая, учитывая огромные числа, в большинстве случаев оказывалась «статистически значимой». Когда данные основаны на меньших числах и появляется ошибка выборки, результаты становятся еще более нерегулярными (иллюстративные данные можно найти в таблице B1 в работе Raven (2000b) [63]).

Важность придания больших весов наиболее надежным данным при построении этих графиков, можно усилить, отметив, что, если у нас есть 50 человек в возрастной категории, то 5-й и 95-й перцентили представляют собой баллы, лежащие посередине между баллами, полученными 2-м и 3-м индивидом в соответствующих хвостах распределения соответственно. Поэтому сами по себе эти оценки крайне ненадежны.

Очень важные выводы, которые следует сделать из этого отступления, состоят в том, что представление не сглаженных исходных данных может привести (i) к отвлекающим выяснениям причин случайных колебаний и (ii) к серьезно вводящим в заблуждение оценкам при использовании таблиц справочных данных, по которым можно просматривать оценки отдельных лиц или экспериментальных групп.

Таблица 8.1 (выше) показывает (сглаженные) исходные

ные баллы, соответствующие 5-му, 10-му, 25-му, 50-му, 75-му, 90-му и 95-му перцентилем для каждой возрастной группы в Британской стандартизации 1979 года.

Когда данные по Великобритании за 1979 год были разбиты по ряду демографических переменных (регион, социально-экономический статус и т. д.) и проанализированы с использованием методов множественной регрессии (Raven, 1981; 2000b) [56, 63], выяснилось, что 2,6% от дисперсии СПМР обеспечивались регионами. Однако, когда эффект социально-экономического статуса был частично исключен, этот процент дисперсии упал до 0,5%. Таким образом, оказалось, что региональные различия сами по себе не имеют большого значения.

На долю социально-экономического статуса (СЭС) приходилось 4,8% дисперсии. Однако, поскольку на возраст приходилось 46% дисперсии, на СЭС приходилось 8,9% дисперсии, не связанной с возрастом. Это эквивалентно возрастной корреляции между социально-экономическим статусом и СПМР в 0,30.

Таким образом, баланс населения, оцениваемый с помощью СЭС, необходимо принимать во внимание при сравнении одного набора результатов с другим или при попытке обобщения вывода с одной популяции на другую.

Оценка корреляции СПМР с возрастом составила 0,68. Таким образом, более половины отклонений не было «объяснено» возрастом. Следовательно, вывод, что тесты просто измеряют «интеллектуальную зрелость» неверен.

Как и при стандартизации 1938 года, анализ заданий на основе характеристической кривой заданий (ICC-кривая из Теории ответов на тестовые задания (IRT) или кривая Раша) проводился отдельно для каждой социально-экономической и возрастной группы. Хотя подробные цифры занимают слишком много места, чтобы их здесь можно было представить, можно отметить (а), что ICC для отдельных заданий были удивительно похожи на те, которые были опубликованы 40 лет назад, и (б) что, как видно из представленных сводных данных в Таблице 8.2, задания масштабируются примерно одинаково для детей разного происхождения. Совсем недавно Л. Водегель-Матцен (Vodegel-Matzen, 1994) [92] показала, что создание более «реалистичных» заданий (т. е. использование шляп, бананов и т. д. вместо абстрактных фигур) при сохранении логики предметов делает задания проще для всех, но не меняет ни порядок заданий (по трудности), ни порядок прошедших тест индивидов.

Вывод ясен и жизненно важен: невозможно объяснить различия в средних оценках этих групп на том основании, что в каком-то общем смысле тест «чужд образу мышления детей определенного происхождения». За некоторыми важными групповыми и индивидуальными исключениями, которые здесь не обсуждаются, тест генерирует упорядоченные данные, которые должны иметь какое-то значение. Различия между группами нельзя сбрасывать со счетов как «незначимые». Они служат исследования и объяснения.

Стандартизация США в 1980-е годы

В период с 1983 по 1989 год было проведено около 50 исследований по нормированию тестов ПМР в школьных округах Соединенных Штатов Америки (J. Raven, 2000a) [62]. В каждом округе выборка была, насколько это возможно, репрезентативной для данного округа. Конкретная использованная процедура выборки варьировалась от округа к округу, но по причинам, обсуждаемым в Приложении В, ни в одном случае не применялись процедуры квотной выборки. (Процедура выборки, принятая в каждом районе, описана в ранее упомянутых публикациях.) Всего было протестировано более 60 000 учащихся в возрасте от 5 до 18 лет.

Полученные нормы заметно различались между разными школьными округами, а внутри округов - между социально-экономическими и этническими группами.

Как показано в таблицах 8.3 и 8.4, этническая при-

надлежность и социально-экономический статус, видимому, внесли независимый вклад в различия внутри округов. (Важно отметить, что применение методов множественной регрессии не является строго законным не только потому, что распределения не являются гауссовыми, но и потому, что независимые переменные коррелированы друг с другом.)

Таблица 8.2. Стандартные прогрессивные матрицы Равена

1979 Британская стандартизация: корреляции между трудностью заданий, рассчитываемых отдельно для молодых людей из разных социально-экономических слоев (десятичные точки опущены и округлены до двух знаков после запятой)

SES	1 (High)	2	3	4	5	6	7	8 (Low)
1 (High)								
2	99							
3	99	99						
4	98	99	99					
5	97	98	99	99				
6	98	99	99	99	99			
7	95	96	98	98	99	98		
8 (Low)	95	96	98	98	99	99	99	

Соответственно, как показано в таблице 8.4, регрессии проводились дважды с независимыми переменными, вводимыми в разном порядке: один раз с исключением параметра СЭС, а второй раз с исключением параметра этнической принадлежности. Каким бы способом ни вводились переменные, остается расхождение, которое нужно объяснить другим.

Различия между нормами школьных систем, обслуживающих белых учащихся с разным социально-экономическим статусом, были столь же велики, как и этнические различия в школьных округах.

Различия между нормами школьных систем, обслуживающих разные пропорции белых, черных и коренных американцев, похоже, соответствуют различиям в опубликованных статистических данных о весе при рождении, младенческой смертности и частоте серьезных детских заболеваний, опубликованной Бюро США по переписи населения (United States Government, Bureau of the Census, 1984 г.) [90].

Чтобы сделать процесс законным, в ряде школьных округов, в которых было достаточно учеников разной этнической принадлежности, анализ заданий проводился отдельно среди разных этнических групп. Один довольно типичный пример результата показан в таблице 8.5.

Из подобных результатов (которые дублируют результаты, опубликованные А. Дженсеном (Jensen, 1974) [35]) следует, что тест работает одинаково, то есть измеряет одно и то же качество, в каждой группе.

Кроме того, как показано на рисунке 8.2, Д. Хофман (Hoffman, 1983, 1990) [28, 29] продемонстрировал, что линии регрессии ПМР по различным типам академической успеваемости для разных этнических групп были (независимо от целей исследования) практически параллельными, хотя и имели разные точки пересечения с осями координат (линии регрессии, показанные на рисунке 8.2., для теста по математике расходятся, в то время как для теста по чтению сходятся, это всего лишь два примера).

В целом, некоторые расходятся, а некоторые сходятся столь незначительно, что можно сделать общий вывод состоит о практической параллельности.)

Таким образом, хотя этнические группы набирают разные итоговые баллы как по тестам достижений, так и по матричным тестам, тест ПМР имеет одинаковую прогностическую достоверность в каждой группе.

Аналогичные результаты были снова получены А. Дженсеном в 1974 г. (Jensen, 1974) [35].

Таблица 8.3. Стандартные прогрессивные матрицы Подростковые перцентильные нормы 1986 года для этнических групп в Весттауне (США) в контексте британских данных 1979 г. (сглаженных)

Percentile		Age in Years								
		12½	13	13½	14	14½	15	15½	16	16½
95	UK	53	54	54	55	56	57	57	—	—
	Anglo	51	52	53	54	55	56	57	58	59
	Asian	53	54	54	54	55	55	56	57	57
	Hisp	48	49	49	50	51	52	53	53	53
	Black	47	48	49	50	51	52	53	54	54
90	UK	51	52	53	54	54	55	55	—	—
	Anglo	50	51	52	53	54	55	56	57	57
	Asian	50	51	51	52	53	53	54	55	55
	Hisp	45	46	47	48	49	50	51	52	52
	Black	45	46	47	48	49	50	51	52	52
75	UK	47	49	49	50	50	51	51	—	—
	Anglo	46	47	47	48	50	52	53	54	54
	Asian	46	47	48	48	49	50	50	51	52
	Hisp	42	43	44	45	46	47	48	49	50
	Black	42	42	42	42	44	45	46	49	49
50	UK	42	43	44	45	46	47	47	—	—
	Anglo	41	42	43	44	47	48	48	48	49
	Asian	42	43	43	43	44	45	46	47	48
	Hisp	37	38	39	40	41	42	43	44	45
	Black	36	36	37	38	39	40	41	43	44
25	UK	38	39	41	42	42	42	42	—	—
	Anglo	37	38	39	40	42	44	45	45	45
	Asian	35	35	36	36	37	38	40	42	43
	Hisp	32	33	34	35	36	37	39	39	40
	Black	29	30	31	32	33	34	35	36	37
10	UK	32	33	35	36	36	36	36	—	—
	Anglo	32	33	34	35	36	38	40	40	40
	Asian	24	25	26	27	29	30	31	32	33
	Hisp	24	25	26	27	28	29	30	31	32
	Black	25	26	27	28	29	30	31	32	33
5	UK	27	28	29	30	33	33	33	—	—
	Anglo	27	28	29	30	32	34	36	37	38
	Asian	17	18	19	20	23	25	26	28	29
	Hisp	20	21	22	23	23	23	24	25	26
	Black	12	15	17	19	21	23	25	26	26
n (unweighted)	UK	174	185	180	196	189	191	171	—	—
	Anglo	46	59	44	52	53	36	56	40	49
	Asian	31	42	47	48	48	38	55	27	55
	Hisp	35	44	52	45	52	35	48	34	45
	Black	37	57	54	53	39	45	48	42	47

Note. The town name "Westtown" was chosen, at the request of the school district, to preserve anonymity.

Примечание. Название города «Весттаун» было выбрано по требованию школьного округа для сохранения анонимности.

Исследования со Словарной шкалой Милл-Хилл

Пока что мы рассматривали исследования с ПМР. Теперь перейдем к тесту МХС (MHV).

Британская стандартизация 1979 г.

Словарная шкала Милл-Хилл (МХС) была стандартизирована вместе со Стандартными ПМР в исследовании 1979 года среди молодых людей, выборка для которого уже была описана. Как и в случае с СПМР, не осталось различий с показателями МХС в зависимости от региона после того, как эффект социально-экономического статуса был частично исключен (см. Raven 1981; 2000b) [58, 63]. Параметр СЭС объяснял 16,2% дисперсии, не связанной с возрастом. Следовательно, оценки МХС больше связаны с базовым СЭС, чем с оценками СПМР. Параметр Возраст обеспечивал 58% дисперсии МХС. Показатели МХС не выходили на плато в подростковом возрасте так же, как показатели СПМР; рост продолжался примерно на полтора слова за шестимесячный интервал до возраста 15 лет.

Как и в случае с ПМР, анализ отдельных заданий проводился в восьми группах СЭС. Воспроизводимость свойств МХС по группам снова была очень высокой, в среднем 0,97. Таким образом, на порядок, в котором дети приобретают знания о значении слов, домашнее окружение влияет не больше (и не меньше), чем на их способность решать матричные задачи. Представляется неверным, что дети из разных слоев общества изучают разные подмножества словарных слов.

Таблица 8.4. Стандартные прогрессивные матрицы
Данные США за 1986 год для подростков в Вестане.
Вклад этнической принадлежности и социально-экономического статуса в общую дисперсию

	Simple R	Mult. R	R sq.	R sq. change	Beta	Beta sq.
Age	29	29	8	8	27	7
Father's SES	-31	41	16	8	-20	4
Black	24	46	21	5	26	7
Hispanic	14	48	23	2	15	2
Asian	-04	48	23	0	0	0
Age	29	29	8	8	27	7
Black	24	38	14	6	26	7
Hispanic	14	44	19	5	15	2

Таблица 8.5. Стандартные прогрессивные матрицы
Корреляции между трудностью заданий, рассчитываемых отдельно в специфических группах (десятичная точка опущена, значения округлены до двух знаков после запятой)

	Westown Black	Westown White	Westown Hispanic	Westown Asian	Westown All	Des Moines	China
Westown Black							
White	98						
Hispanic	100	98					
Asian	98	99	98				
All	99	99	100	99			
Des Moines	99	97	99	97	99		
China	95	94	94	96	95	96	
UK 1979	99	97	99	98	99	99	97

Стандартизация в США в середине 80-х гг.

Многие школьные округа США, собравшие нормы для ПМР, не применяли МХС. Тем не менее, как видно из Таблицы 8.6, общие нормы для школьников США, рассчитанные на основе накопленных данных, снова отстали от международных показателей. Однако нормы для «белых» в США снова довольно близко соответствовали нормам, достигнутым для других культур. Как и в случае с ПМР, что видно из Таблицы 8.7, тест масштабировался примерно одинаково для (англоговорящих) учащихся из разных социально-экономических слоёв и разного этнического происхождения: оказывается, учащиеся из некоторых слоев общества не выучивают многие из слов, которые включены в МХС, но неизвестны другим культурным группам.

Исследования взрослых

Стандартизация в Великобритании в середине 1940-х гг.

Нормы для взрослого населения Великобритании для СПМР, которые были опубликованы в конце 1940-х годов и которые сформировали основные справочные данные, используемые во всем мире на протяжении более полувека, были получены в результате ряда исследований, проведенных в период с 1939 по 1947 год. Каждый из этих примеров и способы получения и консолидации результатов, подробно описаны Дж. Равеном (Raven, 2000b) [63]. На рисунке 8.3 показаны полученные нормы в виде графика.

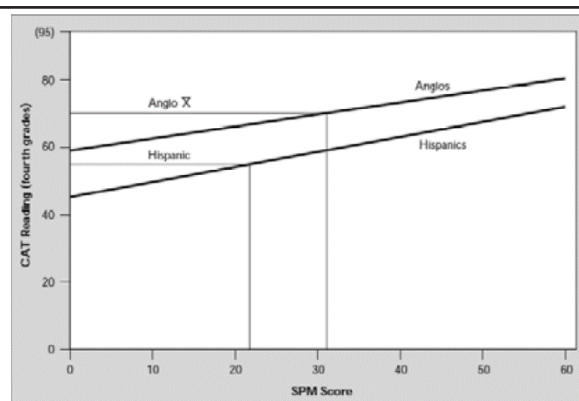
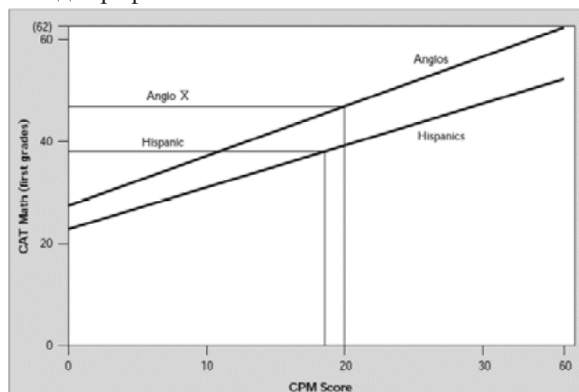


Рисунок 8.2. Цветные и стандартные прогрессивные матрицы

Пример регрессии субтестов Калифорнийского Теста достижений по ПМР для англоязычных и испаноязычных американцев в Дугласе, Аризона.

Примечание. Верхний рисунок показывает регрессию оценок CAT (Cognitive Abilities Test – Тест когнитивных способностей) по математике и по Цветным прогрессивным матрицам Равена среди учащихся первого класса. На нижнем рисунке показана регрессия оценок CAT для чтения и по Стандартным ПМР среди учеников четвертого класса. (Представлено из (Hoffman, 1990) [29].

Хотя данные фактически собирались в течение нескольких лет, для удобства в дальнейшем они будут называться «Британские нормы для взрослых, 1942 г.».

Таблица 8.6. Словарная шкала Милл-Хилл

Подростковые перцентильные Нормы 1986 года для США в контексте британских данных 1979 г. (сглаженных)

Table 8.6. Mill Hill Vocabulary Scale
1986 Adolescent Percentile Norms for the U.S.A. in the Context of 1979 British Data (Smoothed)

Percentile	Age in years (months)															
	6(3)	7(2)	7(3)	8(3)	8(3)	9(3)	9(3)	10(3)	10(3)	11(2)	11(2)	12(2)	12(2)	13(2)	13(2)	14(2)
95	23	22	24	23	26	25	28	27	30	29	32	31	34	33	36	35
90	21	20	22	21	24	23	26	25	28	27	30	29	32	31	34	33
85	18	17	19	18	20	19	22	20	24	22	26	24	28	26	30	28
80	13	12	14	13	15	14	18	16	20	18	22	20	24	22	26	23
75	8	7	9	8	10	9	13	11	14	12	16	14	18	16	20	17
70	6	5	7	6	8	7	10	9	11	10	13	12	15	14	17	15
65	5	4	6	5	7	6	9	8	10	9	12	11	14	13	16	14
60	4	3	5	4	6	5	8	7	9	8	11	10	13	12	15	13
55	3	2	4	3	5	4	7	6	8	7	10	9	12	11	14	12
50	2	1	3	2	4	3	6	5	7	6	9	8	11	10	13	11
45	1	0	2	1	3	2	5	4	6	5	8	7	10	9	12	10
40	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
35	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
30	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
25	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
20	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
15	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
10	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
5	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9
0	0	0	1	0	2	1	4	3	5	4	7	6	9	8	11	9

Note. US figures estimated on the basis of data available Summer 1986. The studies on which these norms are based are detailed in Raven et al. (1990). These show that the norms vary considerably between school districts, and, within districts, between ethnic groups.

Примечание: Американские результаты представлены на основе исследований лета 1986 г. Исследования на которых основаны эти нормы подробно описаны в Raven et al. (1990). Эти результаты показывают, что нормы значительно различаются между школьными округами, а внутри округов – между этническими группами.

Видно, что баллы, полученные в последующих возрастных группах (после 20 лет), становятся все ниже. Однако больше всего «падают» оценки менее способных. Однако следует подчеркнуть, что данные по всем возрастным группам, по сути, собирались одновременно.

Графики не основаны на данных, собранных у тех же людей, которые стали старше. Другими словами, они основаны на поперечных сечениях, а не на лонгитюдных данных.

Как мы увидим, интерпретация данных, как если бы они были лонгитюдными данными, привела к серьезным недоразумениям.

Таблица 8.7. Словарная шкала Милл-Хилл
Корреляции между трудностями заданий, рассчитываемых отдельно в специфических группах (десятичная точка опущена, значения округлены до двух знаков после запятой)

	Westown Black	Westown White	Westown Hispanic	Westown Asian
Westown Black				
Westown White	97			
Westown Hisp.	100	97		
Westown Asian	99	97	99	

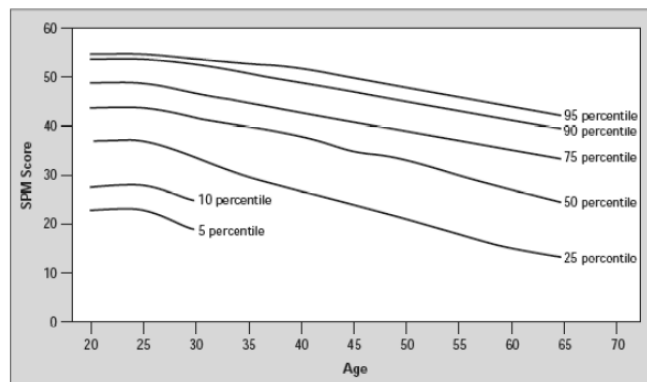


Рисунок 8.3. Стандартные прогрессивные матрицы Равена

Видимое снижение баллов по мере увеличения возраста, что подтверждается типичными перекрестными исследованиями

Стандартизация Великобритании, около 1942 г.

Примечание. На графиках представлены нормы для возрастных срезов людей старшего возраста из всех протестированных в серии исследований, проведенных примерно в 1942 году (по описанию Дж. Равена (Raven, 2000a) [62]).

1992 Стандартизация в Дамфрисе, Шотландия

Первая повторная стандартизация СПМР в Великобритании (которая была объединена с первой в истории стандартизацией Расширенных прогрессивных матриц Равена для взрослого населения в целом) была проведена в Дамфрисе, Шотландия, в 1992 году. Дамфрис был выбран, потому что: (а) исследования, проведенные среди молодежи, описанные ранее, показали, что наиболее важной переменной, которую следует учитывать при выборе места проживания, является баланс групп СЭС в популяции, (б) эти же исследования показали, что нормы ПМР, полученные для округа Бордерс в Шотландии (округа с демографической структурой, соответствующей демографической структуре Великобритании в целом) действительно соответствовали нормам для Великобритании в целом, и (с) город Дамфрис зарекомендовал себя как возможный объект для стандартизации для взрослых, потому что (i) большая часть данных, собранных с помощью ПМРМ за последние 50 лет (включая две основные стандартизации Цветных ПМР), была собрана там и хорошо зарекомендовала себя по сравнению с данными, собранными в других местах, (ii) он имел демографическую структуру, приближенную к британской в целом, и (iii) географически оно было такого размера, что его легко можно было проехать в поисках взрослых по именам, выбранным систематической процедурой отбора объектов из полного списка взрослых жителей. Процедуры, используемые для отбора, установления контакта и тестирования респондентов, вместе с показателями ответов, довольно подробно описаны Дж. Равеном в работе (Raven, 2000b) [63].

1993 Стандартизация в Де-Мойне, Айова

После успеха исследования Дамфриса полностью

аналогичное исследование было проведено в Де-Мойне, штат Айова. Де-Мойн признан одним из четырех городов США с демографическим составом, приближающимся к США в целом, и поэтому широко используется исследователями, ищущими микромир Америки (American Demographics, May, 1985, стр. 38-42). Хотя, конечно, невозможно для какого-либо одного города, сколь бы его приближительная демографическая статистика не соответствовала всей стране, сопоставить все на детальном уровне, исследования, проведенные с помощью ПМР среди школьников в США в 1980-х годах фактически подтвердили, что нормы для Де-Мойна действительно приблизились к нормам для США в целом (Raven, 1989) [60]. Еще раз повторим, процедуры, использованные для составления выборки и установления контакта с потенциальными респондентами (вместе с коэффициентами ответов), описаны в (Raven, 2000b) [63].

Сравнение данных Дамфриса и Де-Мойна

Нормы для тестов «Стандартные ПМР» и «Расширенные ПМР», полученные в Дамфрисе и Де-Мойне, сравнивались друг с другом.

Нормы «Расширенных ПМР» для Де-Мойна сравниваются с нормами для Дамфриса в таблице 8.8. Можно видеть, что нормы для взрослых в Де-Мойне намного хуже, чем можно было бы ожидать в результате нашего предыдущего исследования с большим количеством школьников. Верхние перцентили для Де-Мойна очень близки к данным, полученным в Великобритании, в то время как 50-й и более низкие перцентили (особенно последний) отстают, по крайней мере, до 50 лет. Тем не менее, наши обширные исследования среди молодежи (J. Raven, 2000a; J. Raven & Court, 1989) [62] действительно позволяют предположить, что нижние перцентили для США в целом должны отставать от норм Великобритании в большей степени, чем полученные в Де-Мойне.

Ряд возможных объяснений этого обсуждается в работе (J. Raven et al., 2000, обновлено в 2004) [69], но независимо от объяснения, главный вывод, который здесь следует сделать, состоит в том, что эти нормы Де-Мойна, вероятно, выше тех, которые были бы получены, если бы была протестирована случайная выборка всего населения США.

Таблица 8.9 показывает, что нормы МХС для взрослых в Дамфрисе и Де-Мойне также были сходными.

Влияние даты рождения

Теперь настало время еще раз взглянуть на изменение результатов во времени. Чтобы предвидеть результат, мы обнаружим, что заново интерпретируем данные, полученные из ранее упомянутых перекрестных исследований, которые, как считалось, показывают значительное снижение эдуктивной способности, но меньшее снижение репродуктивной способности с увеличением возраста.

Наш вывод заключается в следующем: более точная интерпретация этих данных состоит в том, что большинство человеческих способностей (включая, например, спортивные способности), но не репродуктивная способность, значительно улучшились за последнее столетие, и это свидетельствует о неожиданном ранее влиянии окружающей среды на результаты тестирования. И все же загадка не в том, в чем принято её видеть. Поскольку большинство способностей улучшаются, возникает вопрос, почему репродуктивные способности меняются так мало, несмотря на инвестиции в образование и средства массовой информации.

Эдуктивная (продуктивная) способность.

На рис. 8.4 показаны нормативные данные 1979 г. для Стандартных ПМР, полученные из ранее описанного общенационального исследования молодых людей в Великобритании (представленного в таблице 8.1) в контексте данных, полученных в исследовании в Ипсвиче, 1938 г.

Таблица 8.8. Расширенные прогрессивные матрицы. Блок II (Без учета времени)

1993 Перцентили норм для взрослых, Де-Мойн, Айова (США) в контексте данных исследования Дамфрис (Великобритания), 1992.

Table 8.8. Advanced Progressive Matrices, Set II (Un timed)
1993 Adult Percentile Norms for Des Moines, Iowa (U.S.) in the Context of 1992 Dumfries (U.K.) Data

		Age in years															
		20	25	30	35	40	45	50	55	60	65	70					
		18	23	28	33	38	43	48	53	58	63	68+					
Per-		to	to	to	to	to	to	to	to	to	to	to					
centile		22	27	32	37	42	47	52	57	62	67		UK	US	UK	US	
95	33	32	33	32	33	32	32	32	31	31	31	30	31	29	30	27	26
90	31	30	31	30	31	30	30	30	29	28	28	27	27	26	25	22	23
75	27	27	27	27	27	26	26	26	25	24	24	23	25	22	24	21	19
50	22	20	22	20	22	20	19	19	18	17	18	16	16	15	14	13	12
25	17	15	17	15	16	15	15	14	13	14	12	13	11	12	10	7	8
10	12	10	12	10	12	10	10	10	9	8	8	7	7	6	4	2	1
5	9	7	9	7	9	7	7	7	6	5	5	4	4	3	3	2	1
n	58	28	71	53	84	72	69	77	54	121	67	69	54	33	39	36	54

Note. Tests completed at leisure.

Примечание: тестирование проводилось в свободное от работы время

Таблица 8.9. Словарная шкала Милл-Хилл, Формы 1 и 2 (Без учета времени)

1993 Перцентили норм для взрослых, Де-Мойн, Айова (США) в контексте данных исследования Дамфрис (Великобритания), 1992.

Table 8.9. Mill Hill Vocabulary Scale, Forms 1 and 2 (Un timed)
1993 Adult Percentile Norms for Des Moines, Iowa (U.S.) in the Context of 1992 Dumfries (U.K.) Data

		Age in Years															
		20	25	30	35	40	45	50	55	60	65	70					
		18	23	28	33	38	43	48	53	58	63	68+					
		to	to	to	to	to	to	to	to	to	to	to					
Perc-		22	27	32	37	42	47	52	57	62	67		UK	US	UK	US	
95	67	69	69	70	71	71	73	72	75	73	77	74	79	75	81	78	86
90	64	63	66	65	68	66	70	67	72	68	74	69	76	70	78	72	82
75	59	56	61	57	63	58	65	59	67	61	68	63	70	65	71	65	74
50	53	51	55	52	57	53	58	54	60	55	61	57	62	58	63	62	63
25	46	44	48	46	50	47	52	48	54	50	55	51	56	53	56	53	53
10	38	36	42	38	44	40	47	42	49	44	49	46	49	46	49	46	45
5	28	23	32	25	36	27	40	31	43	35	43	37	43	38	41	38	36
n	56	26	69	53	81	70	69	75	53	118	60	68	49	31	38	35	56

Note. Tests completed at leisure.

Примечание: тестирование проводилось в свободное от работы время

Важно подчеркнуть, что графы таблицы не отображают баллы, полученные одними и теми же молодыми людьми по мере их взросления: они показывают перцентильные баллы, полученные в результате поперечного возрастного среза выборки молодых людей разного возраста, которые прошли тестирование в один и тот же год. Если на рисунке 8.4. сравнить графики из выборки 1938 года (жирные линии) с графиками для тех же перцентилей для тестируемых в выборке 1979 года (светлые линии), становится ясно, что уровень, на котором возникает плато оценок в подростковом возрасте заметно увеличился, и что молодые люди получают более высокие баллы в более раннем возрасте. (Так, в случае 5-го перцентиля 10-летние в 1979 году получили оценки, аналогичные тем, которые получили 14-летние в 1938 году.)

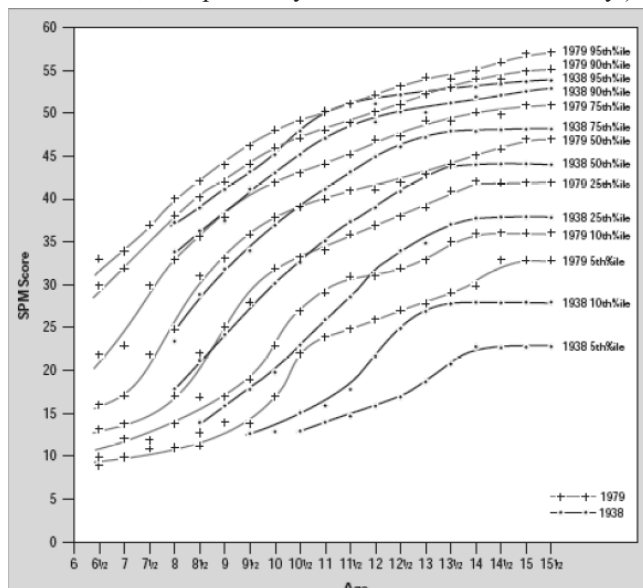


Рисунок 8.4. Стандартные прогрессивные матрицы. Графики процентильных норм для молодых людей в Великобритании в 1938 и 1979 годах.

Примечание. На графиках показаны оценки, полученные молодыми людьми разного возраста и уровня способностей на указанные 2 даты. Если сравнить график норм 1938 года (т. е. жирные линии) с показателями для того же перцентиля в 1979 году (светлые линии), становится ясно, что уровень, на котором возникает плато оценок в подростковом возрасте заметно увеличился, и что молодые люди стали получать более высокие баллы в более раннем возрасте. (Так в случае 5-го перцентиля 10,5-летние подростки в 1979 г. получили оценки, аналогичные тем, которые получили 14-летние в 1938 г.)

Мартинолли (Martinolli, 1990) [39] с Цветными ПМР (1993) [84] и Спичер (Spicher, 1993) [84]) со Стандартными ПМР продемонстрировали аналогичные изменения норм с течением времени во Фрибуре, Швейцария. Касе (Casé) и ее коллеги (см. Raven et. Al. 2000, обновлено в 2004) [69] аналогичным образом задокументировали аналогичные изменения для Аргентины с 1964 по 2000 год).

Эти результаты не показывают, был ли рост непрерывным и постепенным или же он произошел в определенное время, например, во время Второй мировой войны.

Данные Бувье (1969) [7], полученные из ежегодного тестирования призывников в бельгийскую армию с 1958 по 1967 год, и воспроизведенные на рис. 8.5, показывают, что рост был постоянным, а не связанным с какими-либо конкретными событиями.

Результаты СПМР по стандартизации взрослых в 1992 г. в Дамфрисе показаны пунктирными линиями на рис. 8.6 с разбивкой по датам рождения. Пунктирными линиями повторно показаны оценки, полученные в исследовании 1940-х годов, которые были ранее показаны на Рисунке 8.3.

Изучение точек сопряжения двух наборов графиков (например, среди людей, родившихся в 1922 году, где более ранние данные особенно сильны) показывает, что как среднее, так и разброс оценок были очень похожими, независимо от того, были ли они получены из проверенной выборки, в начале 1940-х годов (когда им было примерно 20 лет) или из результатов тестирования в Дамфрисе в 1992 году (когда им было 70 лет). Вместо того, чтобы показывать снижение оценок с возрастом (что, как считалось ранее, демонстрируют данные, стоящие за каждым из этих наборов графиков, а также другие подобные данные), на рисунке ясно показано регулярное и непрерывное увеличение итогового балла, полученного людьми, рожденными в разные годы, при этом баллы более молодых и способных респондентов являются максимально возможными в проводимом тесте.

Как отмечалось выше, непрерывность графиков, полученных для двух выборок, протестированных в разных условиях в разных местах, дает уверенность в адекватности данных, полученных в обоих исследованиях. Эта уверенность усиливается, когда вводятся данные третьего, меньшего по размеру, исследования, проведенного А. Хероном и Ш. Чоун (Heron A, Chown S, 1967) [26] примерно на полпути между двумя уже упомянутыми исследованиями. Данные были наложены на графики, показанные на Рисунке 8.6 на Рисунке 8.7. Графики для данных Херона и Чоун проходят прямо через точку сопряжения между графиками для данных 1942 и 1992 годов на предыдущем рисунке. Таким образом, они подтверждают адекватность данных обоих ранее упомянутых исследований.

В целях защиты от вводящих в заблуждение выводов, сделанных на основе рисунков 8.6 и 8.7, и поскольку эти рисунки на первый взгляд кажутся подтверждающими

вывод о «отсутствии доказательств достижений более высоких уровней», который сделали Тисдейл и Оуэн (Teasdale and Owen, 1989) [86], то важно отметить, что относительно небольшое увеличение итоговых баллов среди более способных людей, родившихся между 1922 и 1972 годами, полностью связано с эффектом потолка для СПМР, который насчитывает всего 60 пунктов. Как уже упоминалось, и в стандартизации Дамфриса, и в Де-Мойне, «Расширенные ПМР» были стандартизированы вместе с СПМР. Нормы «Расширенных ПМР» для Дамфриса на 1992 год сравниваются с нормами 1962 года для того же теста в Таблице 8.10. Сразу очевидно, что увеличение баллов, видимое в нижних перцентилях на рис. 8.6, сопровождалось значительным успехом среди более способных. (Огромные методологические трудности, связанные с любой попыткой выделить относительный размер выигрыша в различных точках, даже в масштабах, соответствующих модели Раша, обсуждаются в главе Прилера и Равена (Prieler, Raven) в этой книге.)

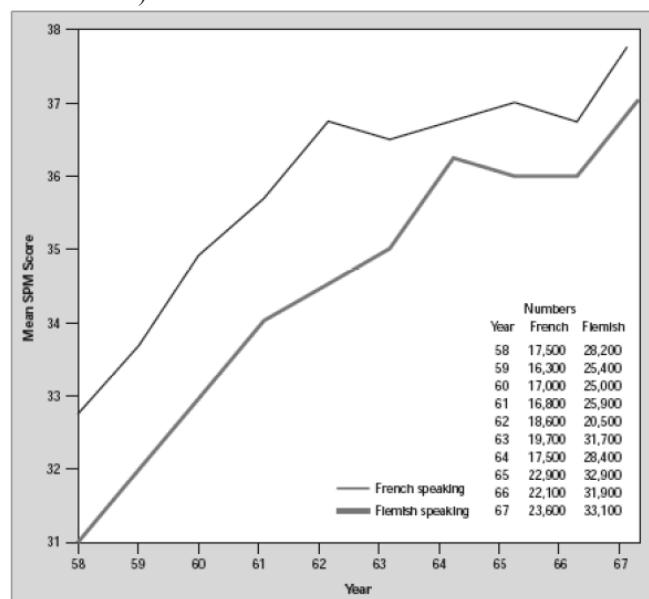


Рисунок 8.5. Стандартные прогрессивные матрицы. Средние итоговые баллы бельгийских призывников с 1958 по 1967 год.

Примечание. Средние итоговые баллы новобранцев, говорящих на французском и фламандском языках, показаны отдельно. (Представлено из (Bouvier, 1969) [7], с разрешения автора.)

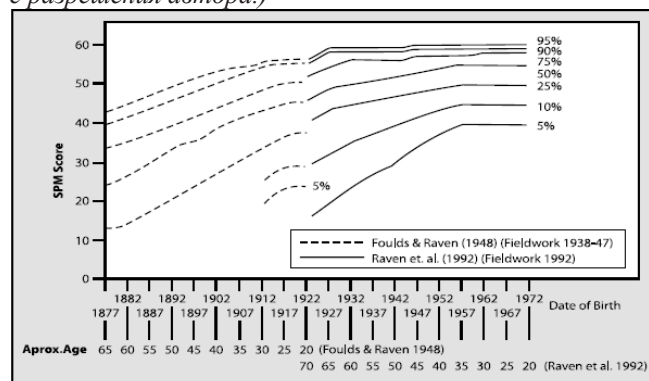


Рисунок 8.6. Стандартные прогрессивные матрицы. 100 лет эдуктивной способности в Великобритании. Графические перцентильные нормы из стандартизаций 1942 и 1992 годов, сопоставленные с датой рождения.

Примечание. На рисунке показаны перцентильные нормы, полученные взрослыми разного возраста (и, следовательно, с разными датами рождения) на Стандартных ПМР, когда выборка была протестирована примерно в 1942 году (см. Комментарий к рисунку

8.3) в одном случае и в 1992 году в другом. Примерный возраст людей, родившихся в разные годы в двух выборках, показан ниже.

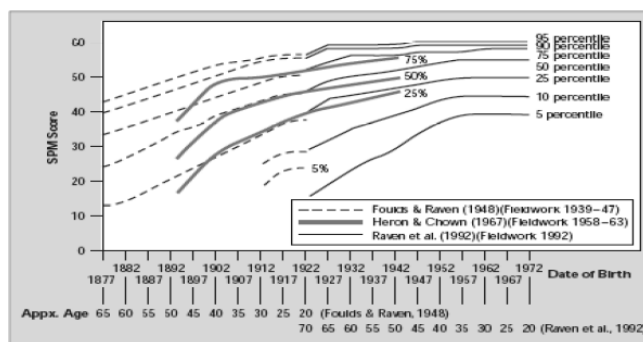


Рисунок 8.7. Стандартные прогрессивные матрицы. Стандартизация Великобритании, около 1942, 1962 и 1992 годов

100 лет эдуктивной способности, включая данные Херона и Чоун (Heron & Chown) за 1962 год [26].

Таблица 8.10. Расширенные прогрессивные матрицы, блок II

Сравнение перцентилей норм взрослого населения Великобритании 1992 и 1962 годов

Percentile	Age in years					
	20		30		40	
	1962	1992	1962	1992	1962	1992
95	24	33	23	33	21	32
90	21	31	20	31	17	30
75	14	27	12	27	9	26
50	9	22	7	22	—	20

Примечание. Данные за 1962 г. (ранее опубликованные в работе (J. C. Raven, 1965) были оценены на основе работы Фулдса и Форбса (Foulds and Forbes), вошедшей в книгу Дж. К. Равена (Raven, J. C. (1965). *Advanced Progressive Matrices, Sets I and II: Plan and Use of the Scale with a Report of Experimental Work Carried out by G. A. Foulds, & A. R. Forbes. London: H. K. Lewis.*). Поскольку тест состоит из 36 заданий и 8 вариантов в каждом задании, оценка б или менее граничит с уровнем случайности. Поэтому в 1965 году не было смысла публиковать нижние перцентили.

Эффект был настолько велик, что тест «Расширенные ПМР», который изначально был разработан для различения верхних 20% населения, теперь предлагает почти идеальное распределение по Гауссу для всего взрослого населения. Подобно тому, как все распределения роста и спортивных способностей переместились вверх (правда, с некоторыми изменениями в форме), то же самое произошло и с полным распределением эдуктивной способности.

Репродуктивная способность

Обратимся теперь к репродуктивной способности. В таблице 8.11 сравниваются британские нормы 1979 года по Словарной шкале Милл-Хилл (МХС) для подростков с нормами, которые были получены с помощью письменного теста в Колчестере в 1943 году. 95-й перцентиль безошибочно упал с уровня 1943 г. до уровня 1979 г. Понижились так же и 90-й, и 75-й перцентили, но их падение менее заметно. 50-й перцентиль практически не изменился; 25-й – поднялся, а 10-й и 5-й перцентили показали заметный рост.

К сожалению, эти явно однозначные результаты не полностью подтверждаются при сравнении результатов 1979 года и результатов, полученных при устном проведении МХС в 1943 году. Возможно, наиболее важно то, что сравнение результатов, полученных при письменном тестировании, предполагает уменьшение дисперсии в 1979 году, сравнение письменного теста 1979 года

и устного теста 1943 года показало увеличение разброса: более способные ученики, как оказалось, знали еще больше, а менее способные ученики знали еще меньше!

Таблица 8.11. Словарная шкала Милл-Хилл: формы 1 и 2 (самозаполнение в письменной форме)

Данные Великобритании Нормы для подростков из стандартизации 1979 г. в контексте 1943 г. Колчестерские данные

Percentile	Age													
	11½	12	12½	13	13½	14	15	16	17	18	19	20	21	22
95	41	42	47	44	50	45	52	46	54	48	57	52	60	56
90	40	39	43	41	47	43	49	45	51	47	53	50	55	53
75	34	35	36	37	40	38	43	40	44	42	45	44	48	47
50	29	31	31	32	33	33	35	35	37	36	38	38	40	41
25	24	25	26	27	27	28	29	30	30	32	31	34	33	36
10	17	21	19	22	21	23	22	25	24	27	25	29	26	31
5	12	16	14	17	16	19	17	21	18	24	19	26	20	28

Примечание. На основе выборок объемом 1419 (данные 1943 г.) и 1304 (данные 1979 г.).

На рисунке 8.8 представлены данные для взрослого населения Великобритании. На графиках показаны перцентильные баллы по Словарной шкале Милл-Хилл, полученные на поперечном срезе взрослых, протестированных в течение 1940-х годов, вместе с результатами, полученными на поперечном возрастном срезе населения в 1992 году. Будет видно, что, хотя, по всей видимости, произошли некоторые изменения (оценки менее способных взрослых среднего возраста кажутся выросшими больше всего) изменения не так велики, как изменения оценок эдуктивной способности.

Данные Бувье (Bouvier, 1969) [7], представленные на рис. 8.9, также показывают незначительные изменения в результатах словарного теста за период его исследования, особенно среди франкоязычной группы.

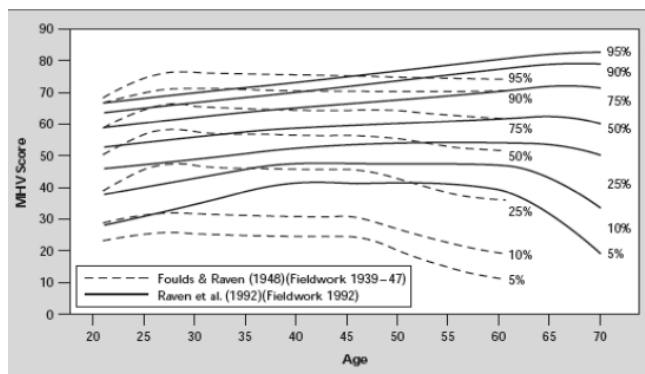


Рисунок 8.8. Стабильность и изменение репродуктивной способности с течением времени по Словарной шкале Милл-Хилл: формы 1 и II

Графики норм из исследований нормирования для возрастных поперечных срезов, проведенных среди британцев

Все эти результаты предполагают, что репродуктивная способность (по крайней мере, оцениваемая на основе знания слов испытуемыми) изменилась гораздо меньше, чем можно было ожидать, и, конечно, намного меньше, чем эдуктивная способность, за период, для которого доступны данные. Шайе (Schaie, 1983, 1994) [75, 76] и Торндайк (Thorndike, 1977) [88] аналогичным образом пришли к выводу, что именно дедуктивные (логические) компоненты «интеллекта» увеличиваются наиболее быстро и последовательно. Их данные особенно интересны тем, что они показывают, что это происходит независимо от того, измеряется ли «способность к рассуждению» («reasoning ability») с помощью вербальных или невербальных тестов и от того, определяется ли репродуктивная способность словарным запасом или дру-

гими рутинными навыками, такими как беглость речи. С другой стороны, их данные действительно предполагают, что знание словаря увеличилось больше, чем можно было бы ожидать, исходя из приведенных выше данных, и что результаты тестов, требующих проявления этих двух способностей в разной степени, увеличились пропорционально степени, в которой они включают эдуктивную способность. Этот вывод был убедительно подтвержден в исследованиях подшкалы WISC (Wechsler Intelligence Scale for Childre - Шкала интеллекта Векслера для детей), проведенных Дж. Флинном (Flynn, 2000) [23].

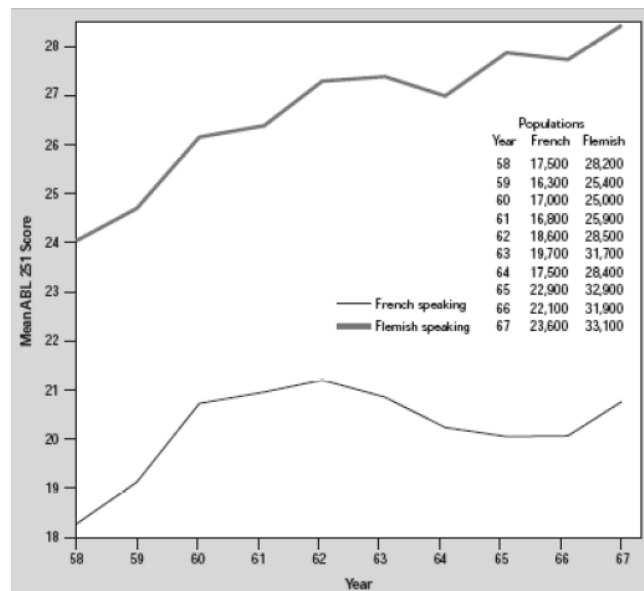


Рисунок 8.9. Словарный тест бельгийской армии (ABL 251)

Средние результаты призывников в бельгийскую армию с 1958 по 1967 год.

Примечание. Средние оценки новобранцев, говорящих на французском и фламандском языках, показаны отдельно. (Представлено по книге Бувье (Bouvier, 1969 г.) с разрешения автора.)

Результаты этих субтестов повысились прямо пропорционально их нагрузке на фактор эдуктивной способности. Позиция К. Шайе по этим вопросам, похоже, с годами изменилась. В одной публикации (Schaie, 1983) [75] он показал, что способность к рассуждению, измеряемая вербальными или невербальными тестами, увеличивается больше всего, а числовая способность увеличивается, а затем уменьшается, а другие способности находятся между ними. В более поздней статье (Schaie, 1994) [76] он представляет графики того, что он называет «когортными градиентами» для «скрытых способностей». Согласно этим данным, средние баллы по «индуктивному мышлению» и «вербальной памяти» с годами увеличились наиболее резко. «Числовые способности» сначала улучшились, а затем пошли на убыль. Его график того, что он называет «вербальной способностью», ведет себя примерно так же, но показывает более поздний пик.

ОБСУЖДЕНИЕ

Из приведенных выше результатов следует, что имело место и остается заметное, хотя и далеко не идеальное, сходство в нормах ПМР, полученных в разных обществах с традициями грамотности в любой данный момент времени. Однако, как и в других тестах, особенно в тестах, измеряющих эдуктивную способность с помощью вербальных или невербальных заданий (Бувье (Bouvier, 1969) [7], Торндайк (Thorndike, 1975, 1977) [87, 88], Гарфинкель и Торндайк (Garfinkel and Thorndike, 1976) [24], а также большое количество опубликованных и неопубликованных исследований, объединенных

вместе Флинном (Flynn, 1984, 1987) [21, 22]) наблюдаются непрерывный рост оценок на всех уровнях способностей с течением времени.

Данные об изменениях во времени и различиях между этническими группами естественным образом поднимают вопрос о том, что вызывает эти изменения и различия. Ни одно исследование, не говоря уже о любом исследовании корреляционного характера, не может дать однозначного ответа. Но, поскольку они, как представляется, делают одни гипотезы менее вероятными, а другие - более вероятными, стоит обобщить некоторые данные, которые имеют отношение к этому вопросу.

В дальнейшем одновременно будут рассматриваться причины изменений во времени и различия между этническими группами. С одной стороны, отсутствие межкультурных различий в итоговых баллах ПМР между культурами, которые действительно различаются по переменной, которая была выдвинута как возможное объяснение изменений во времени, делает эти объяснения временных различий менее вероятными. С другой стороны, различия в оценках между культурными группами, которые действительно различаются по переменной, которая также изменилась с течением времени и была предложена в качестве возможного объяснения временных различий, усиливает возможность того, что эта переменная играет значительную роль и в этом процессе.

Торндайк и Гарфинкель в работах (Thorndike, 1977) [88] и (Garfinkel and Thorndike, 1976) [24] перечислили ряд возможных объяснений временных тенденций. Однако имеющиеся данные о *Прогрессивных Матрицах* не подтверждают их. Торндайк предположил, например, что ускорение развития может быть связано с более ранней зрелостью. Однако, если фактором является зрелость, кривые, изображающие возрастные нормы для мальчиков и девочек отдельно, должны отличаться больше, чем это показывают данные, опубликованные в работах Дж. Равена (J. Raven, 1981) [58] и (J. Raven et al., 2000a) [62]. Из этих данных следует, что, за исключением необъяснимого расхождения между двумя кривыми в возрасте 11 лет (при смене школы с начальной на среднюю), кривые практически идентичны. Кроме того, само это расхождение не было подтверждено, когда мы построили аналогичные графики, например, для ряда школьных округов США. Точно так же Торндайк предположил, что увеличение могло быть связано с изменениями в характере дошкольного образования, но тот факт, что не было большой разницы между нормами ПМР, полученными в Шотландии и Англии в стандартизации 1979 года, показывает, что это маловероятно (потому что шотландское дошкольное образование остается очень формальным (HMI, 1980) [27]). Незначительное различие между китайскими и британскими нормами также опровергает это утверждение. Действительно, некоторые школьные системы, для которых существуют нормы, не принимают детей до достижения ими восьмилетнего возраста, и, как отметил сам Торндайк, наибольшее увеличение итогового балла школьников, по-видимому, произошло среди детей дошкольного возраста. Торндайк предполагает, что телевидение могло оказать влияние. Однако телевидение было широко доступно в Ирландии, когда были собраны данные, которые сейчас считаются низкими ирландскими нормами. Аргумент Гринфилда (Greenfield, 1998) [25] состоящий в том, что это изменение связано с знакомством с иконками и компьютерными играми, также не выдерживает критики, потому что, как показал Шайе (Schaie, 1983) [75], произошло *огромное* увеличение оценок по *вербальным* измерениям «способности к рассуждению» (или эдуктивной способности).

Другие предположили, что увеличение показателей ПМР во времени может быть связано с тем, что школы используют задачи матричного типа для обучения «решению проблем». Однако Торндайк показал, что успеваемость по всем подшкалам Стэнфорд-Бине также улуч-

шилась, и что наибольший прирост был у очень маленьких детей, которые еще не пошли в школу.

По нашим собственным данным, существует небольшая разница между нормами для культур, которые заметно различаются по возрасту, при достижении которого дети идут в школу.

Флинн, подвергший сомнению гипотезу Торндайка относительно результатов Бине в 1984 году, в своей статье 1987 года аналогичным образом пришел к выводу, что большинство общих и очевидных объяснений увеличения итоговых баллов ПМР не срабатывают.

Среди прочего, он показал, посредством подробного анализа данных опубликованных в работе Я. Де Леува и А. Мистера (de Leeuw and Meester, 1984) [16], что изменения в уровне образования, полученного людьми, могут быть причиной только одного из 20-балльных приростов итоговых баллов ПМР, зафиксированных среди военнослужащих. Изменения в интеллектуальном качестве домашней среды - по крайней мере, в той мере, в какой она индексируется социально-экономическим статусом - могут объяснить немногим больше.

Таким образом, большинство распространенных объяснений изменений во времени не актуальны: там, где есть различия между культурами в переменной, которая потенциально помогает объяснить изменение во времени, они не сопровождаются различиями в итоговых баллах ПМР. Таким образом, признавая такие объяснения *менее* вероятными (хотя, конечно, не исключая их), нам следует поискать в другом месте.

На потенциально более плодотворное направление исследования указывает тот факт, что разница в средних показателях между этническими группами в США, действительно соответствует различиям между одними и теми же группами по росту, весу при рождении и младенческой смертности. Рост и вес при рождении, как и результаты тестов на интеллект, увеличились за последние 80 лет (Floud, Wachter, & Gregory, 1990) [20]. Эти наблюдения привели нас к предположению, что увеличение итоговых баллов ПМР с течением времени может быть связано с теми же факторами, которые повлекли увеличение роста и веса при рождении, а также снижение детской смертности, то есть с улучшением питания, повышением благосостояния и гигиены.

Это утверждение, к сожалению, не продвигает нас далеко вперед, так как причины этих изменений остаются неясными. Но то же самое происходит и со способами, которыми большинство лекарств производит свое действие: тот факт, что нельзя предложить полное объяснение, не умаляет ценности того, что можно сделать, чтобы прояснить позицию.

При поиске литературы, подтверждающей гипотезу питания, мы обнаружили замечательное, хотя полностью неопубликованное исследование, проведенное в Абердине, Шотландия, которое показало, что диетические и гигиенические параметры действительно *вливают* на показатели ПМР, а также на вес и рост при рождении (Baird & Scott, 1953; Scott, Illsley, Thomson, 1956) [3, 80]. В этом исследовании потребление кальция использовалось как показатель качества диеты. Это оказало заметное влияние на все три упомянутых показателя, а также на взаимосвязь как внутри, так и между социально-экономическими группами. Ф. Вернон (1969) [91] пришел к аналогичному выводу.

Совсем недавно Бентон и Робертс (Benton and Roberts, 1988) [6], Бентон и Баттс (1990) [4], Бентон и Кук (Benton and Cook, 1991) [5], Нидич, Морхед, Нидич, Сэндс и Шарма (Nidich, Morehead, Nidich, Sands, and Sharma, 1993) [50], Шонталер, Амос, Дораз, и Вэйкфилд (Schoenthaler, Amos, Doraz, Kelly, and Wakefield, 1991) [77], Шонталер, Амос, Айзенк, Перитц и Юдкин (Schoenthaler, Amos, Eysenck, Peritz, and Yudkin, 1991) [78] и другие показали, что витаминные и минеральные добавки оказывают быстрое и заметное влияние на некоторые оценки эдуктивной, но не репродуктивной, спо-

собности людей.

Подтверждение гипотезы о питании/гигиене исходит из обзоров физических факторов, влияющих на интеллект, Поллитт и Сако-Поллитт (Pollitt and Saco-Pollitt, 1996) [54] и Сигман и Уэйли (Sigman and Whaley, 1998) [82]. Низкий уровень йода в рационе питания оказал серьезное влияние на США в прошедшем веке, и все еще имеет место в других странах. Аналогичные эффекты вызывают кишечные паразиты. Низкое содержание кислорода в воздухе, возникающее из-за большой высоты, также оказывает заметное влияние (вспомним ранее упомянутые перуанские горные нормы).

Казалось бы, все эти исследования подтверждают вывод о том, что различия во времени и различия между культурными группами отражают баланс витаминов, в рационе (но не абсолютное количество пищи, как у Р. Марторелла, (Martorell, 1998, [40]) в изолированных горных районах (где вряд ли имеется обширная торговля сельскохозяйственной продукцией), которые имеют самые низкие итоговые баллы. Но этот вывод не получает однозначной поддержки в литературе. Данные Х.К. Чиама особенно сбивают с толку. Это происходит по двум причинам. Во-первых, чтобы определить, что ели испытуемые, Чиама (Chiam, 1995) [12] собрала анализы фекалий у детей из того среза, который демонстрировал различные результаты тестов (но не результаты ПМР). Она обнаружила, что диета не имела значения. В другом исследовании (Chiam, 1994) [11] с участием 5412 детей в возрасте от 7 до 12 лет в контрастных районах Малайзии она обнаружила, что, как показано на Рисунке 8.10, нормы ПМР для китайцев в Малайзии соответствуют международным нормам, в то время как нормы для малайского населения – не соответствуют. Если бы различия были обусловлены питанием, можно было бы ожидать, что они уменьшатся, если частично исключить такие параметры, как влияние городского и сельского проживания и социально-экономический статус. Тем не менее, хотя и существовали различия между городскими и сельскими нормами для обеих групп, эта переменная не учитывала различия между группами. Ни одно из различий не объяснялось образованием или социально-экономическим статусом. И снова анализ делает мало вероятными наиболее часто предлагаемые объяснения групповых различий и изменений во времени, но в этом случае он также ставит под сомнение диетическую гипотезу.

Некоторые последствия

Из этого обзора доступной литературы кажется, что большинство наиболее часто предлагаемых объяснений изменений с течением времени (включая те, которые продвигаются такими авторами, как Гринфилд и Уильямс в книге У. Нейссера (Neisser, 1998) [25, 48]) вряд ли будут верными. Даже наиболее вероятная гипотеза диеты/благополучия/гигиены не получает однозначной поддержки. Но какими бы ни были их причины, задокументированные изменения указывают на явное, серьезное и ранее неожиданное влияние окружающей среды на эдуктивную способность.

Из этого следует, что, несмотря на их постоянство в течение долгого времени, пока оценки *всех* групп росли (вспомним также сохранение различий между говорящими на французском и фламандском языках в Бельгии, показанное на рисунках 8.5 и 8.9), различия между этническими группами внутри США нельзя считать неизменными. Конечно, именно *этот* момент, а не какой-то более общий аргумент о природе «интеллекта», пытался установить Дж. Флинн, когда он приступил к документированию и публикации изменений результатов тестов на «интеллект». Этот момент можно подчеркнуть, отметив, что большинство норм ПМР для этнических показателей в США в настоящее время находится между британскими нормами 1938 и 1979 годов, то есть в пределах диапазона, в котором вызванные окружающей средой изменения могли не только быть теоретически

возможными, но и фактически были продемонстрированы.

Другие особенности окружающей среды, которые имеют значение

Эффектов, обсуждаемых выше, недостаточно для объяснения больших временных и культурных различий в результатах тестов, и было бы неуместно представлять здесь подробный обзор соответствующей литературы. Однако был проведен ряд эмпирических исследований факторов, которые увеличивают или уменьшают итоговые баллы ПМР, и некоторые из них стоит упомянуть как контрапункт упрощенным наследственным и дисгеническим (наследование нежелательных свойств) аргументам об «интеллекте». Результаты удивляют многих психологов. Оказалось, что эдуктивная способность легче поддается влиянию соответствующего образовательного и развивающего опыта, чем репродуктивная способность. Однако переменные, влияющие на развитие эдуктивной способности, не являются очевидными культурными и социально-экономическими переменными, которые разделяют общество и которым социологи уделяют так много внимания. Полученная информация *более подвержена* влиянию этих переменных больше, чем способность воспринимать и ясно мыслить, но эти базовые переменные по-прежнему составляют лишь небольшую часть общей дисперсии.

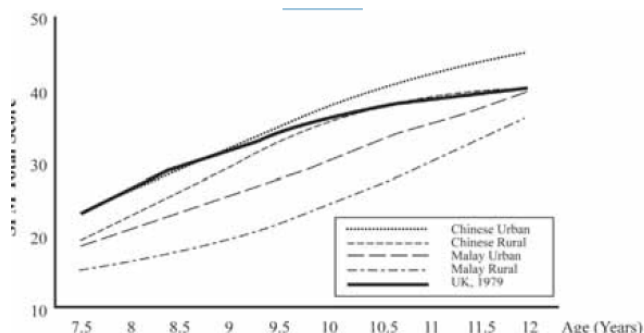


Рисунок 8.10. Стандартные прогрессивные матрицы
Средние результаты молодых людей в Малайзии по
этнической принадлежности и месту жительства
Данные за 1992 год

Примечание. Из исследования Х. Чиама (Chiam, 1994).

Эдуктивная способность, воспитание и образование

Многие исследования (например, Chan, 1981 [9]; McGillicuddy-DeLisi, 1985 [41]; McGillicuddy-DeLisi, DeLisi, Flaughner, & Sigel, 1987 [42]; J. Raven, 1980 [57]; Sigel & Kelley, 1988). [81]) показали, что развитие эдуктивной способности у детей ускоряется, если родители вовлекают их в собственные мыслительные процессы.

Такие родители вовлекают своих детей в собственные попытки разобраться в трудных ситуациях, тем, что они используют свои чувства в качестве основы для «экспериментальных» действий; тем, что они разрешают ценностные конфликты и рассматривают долгосрочные социальные последствия своих действий. Все это требует, чтобы родители делились со своими детьми собственным пониманием работы общества и своей роли в нем. Таким образом, детям предлагается мыслительный процесс, который по своей сути концептуален, но также связывает мысль с действием. Более вероятно, что такие родители будут относиться к своим детям с уважением и осознают необходимость заслужить (а не требовать) уважения с их стороны. Этим инициируется циклический процесс, в ходе которого родители обнаруживают, насколько на самом деле компетентны их дети, и, следовательно, становятся более склонными посвящать их в ситуации, требующие высокой компетентности. В результате их дети получают много возможностей практиковать и развивать такие навыки. Такие родители более

склонны читать своим детям рассказы, затрагивающие моральные проблемы. В результате дети опережают различных персонажей в книгах и могут сформировать свою собственную моральную позицию.

Важность для детей чтения в развитии их моральных качеств и аналогичных рассуждений подчеркивалась в работах Ф. Джексона (1986) [33] и П. Витца (Vitz, 1990) [93].

Дж. Равен (1980, 1987, 1989) [57, 59, 60] и Л.С. Выготский (1978, 1981) [94, 95] показали, что вышесказанное является лишь частью более широкого процесса, посредством которого родители, эффективно воспитывающие высокоуровневые компетентности у своих детей, адаптируют окружающую среду к мотивам, зарождающимся талантам и проблемам своих детей.

Это один из способов, которыми, как показано в работах Р. Пломин (Plomin, 1989) [55] и Р. Пломин и Д. Дэниелс (1987) [56], внутрисемейные различия в детской среде становятся значительными и так связаны с различиями в унаследованных характеристиках, что заметно влияют на развитие детей. Как отмечено в работе [74], аналогичный эффект возникает, когда дети сами попадают в разную среду.

Из этих наблюдений следует, что, если мы хотим идентифицировать генетические и средовые переменные, которые влияют на психологическое развитие, нам потребуется разработать более сложную модель этого процесса.

Развитие эдуктивной способности в школах (но только в некоторых случаях, измеряемых ПМР) изучается многими исследователями.

Никерсон, Перкинс и Смит (Nickerson, Perkins, and Smith, 1985) [51] и Столлинкс и Касковиц (Stallings and Kaskowitz, 1974) [85] обнаружили, что развитию эдуктивной способности способствуют по крайней мере некоторые формы «открытого» или «прогрессивного» образования. Миллер, Кон и Шулер (1 Miller, Kohn, and Schooler, 1985, 1986) [46, 47] и Дж. Равен, Джонстоун и Варли (J. Raven, Johnstone, and Varley, 1985) [64] обнаружили, что самоуправление в образовании (т.е. ученики берут на себя ответственность за собственное образование и моральные решения) и осуществление более сложной образовательной деятельности (например, проектно-ориентированная работа) привели к циклическому развитию когнитивной способности. Большой упор на самоуправление и развитие нового понимания способствует повышению компетентности учащихся, что, в свою очередь, увеличивает желание учащихся обрести больший контроль над своей судьбой и поощряет готовность учителей полагаться на способности своих учеников.

Шулер, Мулату и Месфин (Schooler, Mulatu, Mesfin, 2001) [79] в ходе 30-летнего наблюдения за выборкой, первоначально опрошенной и протестированной в 1964 году, подтвердили свою более раннюю лонгитюдную работу (проведенную в основном с Коном), показывающую, что существенно сложная работа улучшает интеллектуальное функционирование. В замечательном экспериментальном исследовании Ловаглия, Лукас, Хаусер и Марковски (Lovaglia, Lucas, Houser, Thye и Markovsky, 1998) [38] показали, что даже относительно незначительные, экспериментально вызванные изменения воспринимаемого статуса вызывают значительные (до половины стандартного отклонения) изменения в итоговых баллах ПМР. (Стоит заметить, что такие изменения больше, чем обычно достигается путем обучения методам решения проблем.)

Рассмотрев материал, демонстрирующий важность определенных методов развития ребенка и образовательных практик в содействии развитию эдуктивной способности, важно повторить, что ни один из психологических и образовательных процессов, упомянутых выше, не дает эффекта, достаточного для объяснения повышения итоговых баллов ПМР от поколения к поколению.

Кроме того, ни один из видов деятельности, описанных в исследованиях, опубликованных на сегодняшний день, существенно не уменьшает дисперсию внутри социально-экономических групп и внутри семей. Тем не менее, вариации внутри семьи составляют две трети расхождений в результатах тестов. Таким образом, кажется, что факторы окружающей среды, которые имеют наибольшее влияние на эдуктивную способность, не являются психологическими и образовательными переменными, которыми психологи занимались в прошлом, и эти переменные, по-видимому, мало влияют на ее наследование.

В поисках однофакторных объяснений

В Общем введении к этой книге мы утверждали, что свидетельства того, что и ПМР, и МХС соответствуют требованиям теории ответов на тестовые задания (Item Response Theory) показывают, что эдуктивная и репродуктивная способности столь же «реальны» и измеримы, как и «твердость» минералов в геологии, «рост» в физиологии, «прыгучесть в высоту» в легкой атлетике и «продолжительность жизни» в статистическом учете.

Там же мы представили график, показывающий, что ожидаемая продолжительность жизни мужчин в Великобритании почти удвоилась за столетие, для которого были доступны данные ПМР. Мы предположили, что этот график опровергает большинство аргументов, выдвинутых Дж. Флинном в его попытке подорвать значимость психологических тестов. Он утверждал, например, что обратная проекция увеличения результатов ПМР и других тестов «интеллекта» на время древних греков означала бы, что они, должно быть, были невероятно глупыми, и, поскольку этого не могло быть, тесты должны быть признаны бессмысленными. К сожалению, для Флинна, те же аргументы применимы к росту людей, спортивным способностям и продолжительности жизни. (Как будет показано в следующих главах, мы не оспариваем его аргумент о том, что важность этих психологических способностей как детерминанта эффективной жизнедеятельности была преувеличена.)

Мы использовали те же аналогии, чтобы привлечь внимание к нелогичности других аргументов, возникающих при обсуждении природы эдуктивной и репродуктивной способностей и переменных окружающей среды, которые на них влияют.

Никто не стал бы использовать масштабируемость какой-либо из ранее упомянутых переменных в качестве основы для утверждения, что наблюдаемая и измеряемая дисперсия должны быть вызваны единственной основной причиной, как это делали многие психологи, утверждавшие, что поскольку эдуктивная способность масштабируется, то отклонение должно быть связано с одним основным фактором, таким как «скорость нейронной передачи». У них не возникло соблазна утверждать, что, поскольку прыгучесть в высоту, очевидно, можно тренировать, то эта мера бессмысленна. Еще меньше они будут склонны искать единственную причину увеличения роста, высокой прыгучести и ожидаемой продолжительности жизни, которые произошли за эти годы.

Тем не менее, есть один заключительный, очень интригующий и почти полностью упущенный вопрос, который возникает из этого исследования.

Он состоит в следующем. Почему, когда все остальное растет (рост, спортивные способности, ожидаемая продолжительность жизни) репродуктивная способность, измеряемая МХС или любым из множества других мер, существующих в этой области, почти не увеличивается вообще? И это несмотря на огромные инвестиции, сделанные в образование, средства массовой информации и информационные технологии. Получается, что в некотором смысле Дж. Флинн сосредоточил внимание психологов на совершенно неправильном вопросе.

ЗАКЛЮЧЕНИЕ

В заключение представляется целесообразным обратить внимание на серьезность ошибок, связанных с использованием устаревших норм. Во-первых, из рисунка

8.6 очевидно, что оценка, которая помещает 50-летнего человека, прошедшего тестирование в 1942 году, на 95-й перцентиль, если его вместо норм 1942 года оценивать по сегодняшним нормам, приведет к классификации на 25-й перцентиль. Такие огромные расхождения в интерпретации оценок означают, что использование устаревших норм не может быть оправдано: они вредны для заинтересованных лиц, вредны для организаций, их использующих, и вредны для общества.

Однако еще более серьезными являются ошибки, возникающие в результате принятия устаревших норм в исследованиях. Эффективность таких вещей, как образовательные программы повышения квалификации, обычно оценивается путем сравнения результатов, полученных экспериментальными группами, с опубликованными нормами. Когда эти нормы устаревают, такие экспериментальные программы могут только казаться намного более эффективными, чем они есть.

ПРИЛОЖЕНИЕ

Процедуры формирования выборок, размеры выборок и управление данными

В этом Приложении изложены некоторые соображения, которыми мы руководствовались при определении методологии формирования выборок и анализа, и представления данных.

Практически все статистические тесты предполагают, что группы, между которыми желательно провести различие или на основании которых предлагается делать обобщение, являются случайными выборками из более широкой совокупности. Тем не менее, придавая большое значение сложным статистическим методам, психологи редко проверяют качество своих выборок. Они нередко предполагают, например, что результаты, полученные при изучении студентов-психологов, будут применимы ко всем людям или ко всем людям в некоторой категории (например, к мужчинам и женщинам).

Обычно, даже когда прилагаются усилия к тому, чтобы протестированная популяция была репрезентативной для более широкой популяции, используются методы «квотной выборки». При этом делается попытка гарантировать, что демографические характеристики тестируемого населения соответствуют характеристикам некоторой более широкой группы населения, на которую предполагается обобщить результаты исследования.

Тем не менее, даже к тому времени, когда Г. Хайман написал свою классическую книгу по планированию и анализу опросов *Survey Design and Analysis* (Нуман, 1955) [32], было неоднократно продемонстрировано, что не только опросы общественного мнения, основанные на огромных числах, дают гораздо менее точные данные, чем исследования, основанные на гораздо меньших выборках, выбранных случайным образом, но менее точны и обследования на основе квотированных выборок.

По этим причинам в нашей собственной работе мы стремились использовать процедуры систематической случайной выборки, где это возможно, делая это в пределах слоев, выбранных для получения правильных пропорций в определенных демографических категориях, необходимых для соответствия более широкой демографической статистике.

Важно отметить, что стратификация с помощью демографической статистики - это совсем другое дело, чем просить отдельных исследователей найти и протестировать определенное количество людей в пределах ряда категорий, определенных с точки зрения таких показателей, как пол, возраст, социально-экономический статус и этническая группа.

В британском исследовании 1979 г. мы смогли с помощью средств Совета по исследованиям в области социальных наук провести исследование в семи регионах страны, которые, как показало предыдущее исследование Р. Вебера (Webber, 1977) [96], покрывают основное различие в пределах страны, в то же время являясь коллективным представителем страны в целом. Мы даже

смогли провести избыточную выборку по конкретным областям, для получения достаточно большого количества респондентов, чтобы можно было проводить подробные сравнения между областями, а затем повторно взвесить данные для получения правильного эффекта в сочетании с другими данными в общей статистике.

В большинстве других работ, описанных в этой главе, это невозможно. Было необходимо работать с сотрудниками, которые были заинтересованы в участии в исследовании и сделали то, что было возможно в данных обстоятельствах. Насколько это возможно, мы нашли как (1) районы с демографически сбалансированным населением, так и (2) диапазон районов, расположенных в частях страны с очень разными демографическими характеристиками.

Внутри территорий мы постарались обеспечить, чтобы тестируемые выборки были отобраны с использованием строго случайного метода. В некоторых случаях были получены полные списки имен, а затем произведена выборка с использованием случайного начала и фиксированного интервала выборки. В других случаях приходилось идти на компромисс, делая такие вещи, как систематический выбор зданий и классов в школьных округах, которые будут репрезентативными для всего округа, а затем тестировать всех детей в этих классах. Такая кластеризация увеличивает числа, но на самом деле не дает лучших выборок. Естественно, данные, полученные таким образом (в отличие от данных, полученных в результате стандартизации Великобритании в 1979 году среди молодежи) не могут обрабатываться с использованием обычных статистических процедур. Вместо этого их необходимо комбинировать с должным учетом недостатков в наборе данных и придавая больший вес более сбалансированным и полным выборкам.

REFERENCES:

1. Adams, E. A. (1952). *Analysis of Raven's Matrices Scores: Preliminary Report*. Surrey, England: Surrey Educational Research Association.
2. Angelini, A. L., Alves, I. C. B., Custodio, E. M., & Duarte, W. F. (1988). *Manual Matrices Progressivas Coloridas*. Sao Paulo, Brazil: Casa do Psicologo.
3. Baird, D., & Scott, E. M. (1953). *Intelligence and childbearing*. *Eugenics Review*, 45(3), 139-154.
4. Benton, D., & Butts, J. (1990). *Vitamin/mineral supplementation and intelligence*. *The Lancet*, 335, 1158-1160.
5. Benton, D., & Cook, R. (1991). *Vitamin and mineral supplements improve the intelligence scores and concentration of six-year-old children*. *Personality and Individual Differences*, 12, 1151-1158.
6. Benton, D., & Roberts, G. (1988). *Effect of vitamin and mineral supplementation on intelligence of a sample of schoolchildren*. *The Lancet*, 23, January, 140-143.
7. Bouvier, U. (1969). *Evolution des Cotes a Quelques Test*. Belgium: Centre de Recherches, Forces Armees Belges.
8. Byrt, E., & Gill, P. E. (1973). *Standardisation of Raven's Standard Progressive Matrices and Mill Hill Vocabulary for the Irish Population: Ages 6-12*. Unpublished Master's Thesis, National University of Ireland, University College Cork.
9. Chan, J. (1981). *Correlates of parent-child interaction and certain psychological variables among adolescents in Hong Kong*. In J. L. M. Binnie-Dawson (Ed.), *Perspectives in Asian Cross-Cultural Psychology* (pp. 117-131). Lisse, Netherlands: Swets and Zeitlinger.
10. Chan, J. (1989). *The use of Raven's Progressive Matrices in Hong Kong: A critical review*. *Psychological Test Bulletin*, November, 2(2), 40-45. Hawthorn, Victoria: ACER.
11. Chiam, H. K. (1994). *Is the Raven Progressive Matrices valid for Malaysians? Paper presented to the 23rd International Congress of Applied Psychology*, Madrid.
12. Chiam, H. K. (1995). *The standardisation of several tests in Malaysia*. Unpublished manuscript, School of Education, University of Malaya, Kuala Lumpur.
13. Court, J., & Raven, C. J. (2001). *A Researcher's Bibliography for Raven's Progressive Matrices and Mill Hill Vocabulary Scales. Obtainable in hard copy and disk format from Susan Middleton, Harcourt Assessment, 19500 Bulverde Rd., San Antonio, Texas 78259, USA. <susan_middleton@harcourt.com>*
14. Court, J. H., & Raven, J. (1995). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 7: Research and References: Summaries of Normative, Reliability, and Validity Studies and References to all Sections*. San Antonio, TX: Harcourt Assessment.
15. Dahlstrom, W. G. (1993). *Tests: Small samples, large consequences*. *American Psychologist*, 48(4), 393-399.
16. de Leeuw, J., & Meester, A. C. (1984). *Over het intelligente-nderzoek bij de militaire keuringen vanaf 1925 tot henden*. [Intelligence-as test-

ed at selections for the military service from 1925 to the present]. *Mens en Maatschappij*, 59, 5-26.

17. de Lemos, M. M. (1984). A note on the Australian norms for the Standard Progressive Matrices. *Bulletin for Psychologists*, (ACER Melbourne), 36, 9-12.

18. de Lemos, M. M. (1989). The Australian re-standardisation of the Standard Progressive Matrices. *Psychological Test Bulletin*, November, 2(2), 17-24. Hawthorn, Victoria: ACER.

19. Ferjencik, J. (1985). *Manual: Coloured Progressive Matrices*. Bratislava, Slovakia: Psychodiagnostické a Didaktické Testy.

20. Floud, R., Wachter, K., & Gregory, A. (1990). *Height, Health, and History*. Cambridge, England: Cambridge University Press.

21. Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29-51.

22. Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171-191.

23. Flynn, J. R. (2000). IQ gains, WISC subtests and fluid g: g theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The Nature of Intelligence* (Novartis Foundation Symposium 233) pp. 202-227. Chichester, England: Wiley.

24. Garfinkel, R., & Thorndike, R. L. (1976). *Binet item difficulty: Then and now*. *Child Development*, 47, 959-965.

25. Greenfield, P. M. (1998). The cultural Evolution of IQ. In U. Neisser (Ed.), *The Rising Curve* (pp. 81-124). Washington, DC: American Psychological Association.

26. Heron, A., & Chown, S. (1967). *Age and Function*. London: Churchill.

27. HMI (Scotland) (1980). *Learning and Teaching in Primary 4 and Primary 7*. Edinburgh: HMSO.

28. Hoffman, H. V. (1983). *Regression Analysis of Test Bias in the Raven's Progressive Matrices for Anglos and Mexican-Americans*. Unpublished doctoral dissertation, Department of Educational Psychology, Graduate College, University of Arizona.

29. Hoffman, H. V. (1990). In J. Raven et al., *A Compendium of North American Normative and Validity Studies*, Research Supplement No.3 to the *Manual for Raven's Progressive Matrices and Vocabulary Tests* (pp. 21-31). San Antonio, TX: Harcourt Assessment.

30. Hollingshead, A. B. (1967). In C. M. Bonjean, R. J. Hill, & S. D. McLemore (Eds.), *Sociological Measurement: An Inventory of Scales and Indices*. San Francisco: Chandler.

31. Holmes, B. J. (1980). *British Columbia Norms for Wechsler Intelligence Scale for Children-Revised; Peabody Picture Vocabulary Test; Slosson Intelligence Test; Standard Progressive Matrices, and Mill Hill Vocabulary Scale*. Unpublished manuscript, University of British Columbia: Faculty of Education.

32. Hyman, H. H. (1955). *Survey Design and Analysis: Principles, Cases, and Procedures*. Glencoe, IL: Free Press.

33. Jackson, P. W. (1986). *The Practice of Teaching*. New York: Teachers College Press.

34. Javorowska, A., & Szustrowa, T. (1991). *Podrecznik Do Testu Matrycy Ravena*. Warsaw: Pracownia Testow Psychologicznych Polskiego Towarzystwa Psychologicznego.

35. Jensen, A. R. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs*, 90, 185-244.

36. Kahn, H., Spears, J., & Rivera, L. (1977). *Applicability of Raven Progressive Matrices Tests with School Children in Puerto Rico*. Hato Rey, Puerto Rico: Department of Education.

37. Kratzmeier, H., & Horn, R. (1979). *Manual: Raven-Matrizen-Test, Standard Progressive Matrices*. Weinheim, Germany: Beltz Test.

38. Lovaglia, M. J., Lucas, J. W., Houser, J. A., Thye, S. R., & Markovsky, B. (1998). Status processes and mental ability test scores. *American Journal of Sociology*, 1, 195-228.

39. Martinolli, L. (1990). *Etude et Reetallonnage des Matrices Progressives Couleur*. Unpublished manuscript, L'Institut de Psychologie, Université de Fribourg.

40. Martorell, R. (1998). Nutrition and the worldwide rise in IQ scores. In U. Neisser (Ed.), *The Rising Curve*. Washington, DC: American Psychological Association.

41. McGillicuddy-DeLisi, A. V. (1985). The relationship between parental beliefs and children's cognitive level. In I. E. Sigel (Ed.), *Parental Belief Systems: The Psychological Consequences for Children* (pp. 261-299). Hillsdale, NJ: Erlbaum.

42. McGillicuddy-DeLisi, A. V., DeLisi, R., Flaugh, J., & Sigel, I. E. (1987). Family influences on planning. In S. L. Friedman, E. K. Scholnick, & R. R. Cocking (Eds.), *Blueprints for Thinking: The Role of Planning in Cognitive Development* (pp. 395-427). New York: Cambridge University Press.

43. Mehlhorn, H. G. (1980). *Aspekte der Geistigen Entwicklung Jugendlicher*. In W. Freidrich und H. Muller (Eds.), *Zur Psychologie der 12 bis 22 Jahrigen*. Berlin, Germany: VEB Deutscher Verlag der Wissenschaften.

44. Miao, E. S. Y. (1993). Translation of J. Raven, J. C. Raven, & J. H. Court, *Manual for Raven's Progressive Matrices Tests*. Taiwanese Edition. Taiwan: Chinese Behavioural Science Corporation.

45. Miao, E. S. Y., & Huang, W. (1990, July). Standardisation and validation of CPM, SPM, and APM in Taiwan, ROC. Paper presented to 22nd International Congress of Applied Psychology, Kyoto, Japan.

46. Miller, K. A., Kohn, M. L., & Schooler, C. (1985). Educational self-direction and the cognitive functioning of students. *Social Forces*, 63, 923-944.

47. Miller, K. A., Kohn, M. L., & Schooler, C. (1986). Educational self-direction and personality. *American Sociological Review*, 51, 372-390.

48. Neisser, U. (Ed.) (1998). *The Rising Curve*. Washington, DC: American Psychological Association.

49. New Zealand Council for Educational Research. (1984). *Standard Progressive Matrices: New Zealand Norms Supplement*. Wellington, New Zealand: NZCER.

50. Nidich, S. I., Morehead, P., Nidich, R. J., Sands, D., & Sharma, H. (1993). The effect of the Maharishi Student Rasayana Food Supplement on non-verbal intelligence. *Personality and Individual Differences*, 15, 599-602.

51. Nickerson, R., Perkins, D. N., & Smith, E. (1985). *The Teaching of Thinking*. Hillsdale, NJ: Erlbaum.

52. Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences*, 13, 149-159.

53. Owens, W. A. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology*, 57, 311-325.

54. Pollitt, E., & Saco-Pollitt, C. (1996). On the role of the physical environment in the development of intelligence. In D. K. Detterman (Ed.), *Current Topics in Human Intelligence: Volume 5: The Environment*. Norwood, NJ: Ablex.

55. Plomin, R. (1989). Environment and genes. *American Psychologist*, 44(2), 105-111.

56. Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from one another? *Behavioral and Brain Sciences*, 10, 1-15.

57. Raven, J. (1980). *Parents, Teachers and Children: An Evaluation of an Educational Home Visiting Programme*. Edinburgh: Scottish Council for Research in Education.

58. Raven, J. (1981). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.1, The 1979 British Standardisation of the Standard Progressive Matrices and Mill Hill Vocabulary Scales, Together With Comparative Data From Earlier Studies in the U.K., US, Canada, Germany, and Ireland*. San Antonio, TX: Harcourt Assessment.

59. Raven, J. (1987). Values, diversity, and cognitive development. *Teachers College Record*, 89, 21-38.

60. Raven, J. (1989). The Raven Progressive Matrices: A review of national norming studies and ethnic and socioeconomic variation within the United States. *Journal of Educational Measurement*, 26, 1-16.

61. Raven, J. (1995). Methodological problems with the 1992 standardisation of the SPM: A response. *Personality and Individual Differences*, 18(3), 443-445.

62. Raven, J. (2000a). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No.3 (Second Edition): A Compendium of International and North American Normative and Validity Studies Together with a Review of the Use of the RPM in Neuropsychological Assessment by Court, Drebing, & Hughes*. San Antonio, TX: Harcourt Assessment.

63. Raven, J. (2000b). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.

64. Raven, J., Johnstone, J., & Varley, T. (1985). *Opening the Primary Classroom*. Edinburgh: Scottish Council for Research in Education.

65. Raven, J., Raven, J. C., & Court, J. H. (1998a, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 1: General Overview*. San Antonio, TX: Harcourt Assessment.

66. Raven, J., Raven, J. C., & Court, J. H. (1998b). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 2: The Coloured Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

67. Raven, J., Raven, J. C., & Court, J. H. (1998c). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 4: The Advanced Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

68. Raven, J., Raven, J. C., & Court, J. H. (1998d). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 5: The Mill Hill Vocabulary Scale*. San Antonio, TX: Harcourt Assessment.

69. Raven, J., Raven, J. C., & Court, J. H. (2000, updated 2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Section 3: The Standard Progressive Matrices*. San Antonio, TX: Harcourt Assessment.

70. Raven, J. C. (1941). Standardisation of Progressive Matrices, 1938. *British Journal of Medical Psychology*, XIX, Part 1, 137-150.

71. Raven, J. C., Court, J. H., & Raven, J. (1995). *Raven, Matrices Pogresivas (Escala: CPM, SPM, APM)*. Madrid: TEA Ediciones, S.A.

72. Raven, J. C., & Walshaw, J. B. (1944). Vocabulary tests. *British Journal of Medical Psychology*, 20, 185-194.

73. Sahin, N., & Duzen, E. (1994). Turkish Standardization of the Raven's SPM (Ages 6-15). Paper presented at the 23rd International Congress of Applied Psychology, 17-22 July, Madrid, Spain.

74. Scarr, S., Webber, P. L., Weinberg, R. A., & Wittig, M. A. (1981). Personality resemblance among adolescents and their parents in biologically related and adoptive families. *Journal of Personal Social Psychology*, 40, 885-898.

75. Schaie, K. W. (Ed.). (1983). *Longitudinal Studies of Adult Psychological Development*. New York: Guilford Press.

76. Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist*, 49(4), 304-313.

77. Schoenthaler, S. J., Amos, S. P., Doraz, W. E., Kelly, M. A., & Wakefield, J. (1991). Controlled trial of vitamin-mineral supplementation on intelligence and brain function. *Personality and Individual Differences*, 12(4), 343-350.

78. Schoenthaler, S. J., Amos, S. P., Eysenck, H. J., Peritz, E., & Yudkin, J. (1991). Controlled trial of vitamin-mineral supplementation: Effects on intelligence and performance. *Personality and Individual Differences*, 12(4), 351-362.

79. Schooler, C., Mulatu, M. S., & Mesfin, S. (2001). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging* 16(3) 466-482.
80. Scott, E. M., Illsley, R., & Thomson, A. M. (1956). A psychological investigation of primigravidae: Part II: Maternal social class, age, physique and intelligence. *Journal of Obstetrics and Gynaecology of the British Empire*, LXIII(3), 338-343.
81. Sigel, I. E., & Kelley, T. D. (1988). A cognitive developmental approach to questioning. In J. Dillon (Ed.), *Classroom Questioning and Discussion: A Multi-disciplinary Study*. Norwood, NJ: Ablex.
82. Sigman, M., & Whaley, S. E. (1998). The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The Rising Curve*. Washington, DC: American Psychological Association.
83. Snow, R. E., Kyllonen, P. C., & Marshalek, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*, Volume 2 (pp. 47-103). Hillsdale, NJ: Lawrence Erlbaum Associates.
84. Spicher, S. (1993). *Nouvel Etalonnage du SPM-38: Raven dans le ciel Fribourgeois*. Unpublished manuscript, L'Institute de Psychologie, Universite de Fribourg.
85. Stallings, J., & Kaskowitz, D. (1974). *Follow Through Classroom Observation Evaluation 1972-1973*. Menlo Park, CA: Stanford Research Institute. Report URU-7370.
86. Teasdale, T. W., & Owen, D. R. (1989). Continuing secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, 13, 255-262.
87. Thorndike, R. L. (1975). Mr. Binet's Test 70 Years Later. Presidential Address to the American Educational Research Association.
88. Thorndike, R. L. (1977). Causation of Binet IQ decrements. *Journal of Educational Measurement*, 14, 197-202.
89. Tuddenham, R. D., Davis, L., Davison, L., & Schindler, R. (1958). An experimental group version for school children of the Progressive Matrices. Unpublished manuscript, University of California. See also Abstract: *Journal of Consultant Psychology*, 22, 30.
90. United States Government, Bureau of the Census (1984). *Statistical Abstract of the United States, 1983*. Washington, DC: U.S. Government Printing Office.
91. Vernon, P. E. (1969). *Intelligence and Cultural Environment*. London: Methuen.
92. Vodegel-Matzen, L. B. L. (1994). *Performance on Raven's Progressive Matrices*. Ph.D. Thesis, University of Amsterdam.
93. Vitz, P. C. (1990). The use of stories in moral development. *American Psychologist*, 45(6), 709-719.
94. Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.
95. Vygotsky, L. S. (1981). The genesis of higher mental function. In J. V. Wertsch (Ed.), *The Concept of Activity in Society Psychology*. Annank, NH: Sharpe.
96. Webber, R. J. (1977). *The National Classification of Residential Neighbourhoods*. PRAG Technical Paper TP23. London: Centre for Environmental Studies.
97. Zhang, H.-C., & Wang, X.-P. (1989). Chinese standardisation of Raven's Standard Progressive Matrices. *Psychological Test Bulletin*, November, 2(2), 36-39. Hawthorn, Victoria: ACER.

Received date: 17.01.2021

Revised date: 12.03.2021

Accepted date: 27.05.2021