

---

# Новая психометрика и оценка понимания прочитанного

*Е. Л. Григоренко,*

*кандидат психологических наук, доцент кафедры психологии образования и педагогики факультета психологии Московского государственного университета имени М.В. Ломоносова, elenalgrigorenko@gmail.com*

---

В этой статье (кратком сообщении) обсуждаются основные характеристики Many-Facet Rasch Model, (MFRM или многофасетной модели Раша) и приводится иллюстрация применения этой модели на примере исследования индивидуальных различий в процессе понимания прочитанного. Иллюстрация основана на данных от 4020 тестируемых, чьи ответы на вопросы о прочитанном материале были рядоположены на единой латентной шкале способности, несмотря на огромный диапазон возрастных характеристик тестируемых и разнообразие вопросов/заданий.

**Ключевые слова:** процесс понимания прочитанного, классическая и новая психометрика, индексы соответствия Модели Раша, Многофасетные модели

---

Создание психометрически адекватного инструментария, который мог бы использоваться как в исследовательской, так и в практической деятельности, является одной из основных задач психодиагностики и психологии индивидуальных различий. В России, где в этих двух подразделах психологии по-прежнему проявляются последствия Постановления ВКП(б) о педологии [4], отсутствие необходимого специального инструментария часто является камнем преткновения при проведении исследований, сопоставимых по содержанию и составу инструментария с исследованиями в западной психологии.

С начала XXI в. западные психология образования и психология развития стали уделять большое внимание изучению психологического процесса понимания прочитанного [11]. Психологическая сторона понимания прочитанного сложна, этот процесс возникает как бы на стыке развития компонентов чтения, необходимых для декодирования и означивания единичных слов, метакогнитивных процессов – инференциальных процессов, а также внимания и памяти, регулирующих продвижение по связному тексту и его осмысление. Значение процесса понимания прочитанного сложно переоценить, поскольку он является

фундаментом усвоения и накопления академических знаний и, кроме того, в некоторых парадигмах тестирования используется как прокси для IQ. Именно процесс понимания тестируется в таких международных сравнительно-оценочных проектах [1; 5], как PIRLS – Международное исследование прогресса в области грамотности чтения (Progress in International Reading Literacy Study) и PISA – Международная программа по оценке образовательных достижений учащихся (Programme for International Student Assessment).

В контексте подобных проектов, как известно, российские школьники выглядят не очень благополучно. Так, по результатам тестирования PISA в 2000 г. Россия была на 28 месте (из 32 стран, которые приняли участие), в 2003 г. – на 32–34 местах (среди 41 страны-участницы), в 2006 г. – на 37–40 местах (среди 56 стран-участниц), в 2009 г. – на 43 месте (среди 65 стран-участниц). Очевидно, что, для того чтобы понять и, желательно, изменить эту картину, процесс понимания прочитанного надо исследовать. В России, к сожалению, подобных исследований проведено крайне мало [в качестве примера см.: 2; 3]. Одна из причин такого положения – отсутствие инструментария, необходимого для оценки процессов понимания прочитанного.

При разработке тестов на понимание требуются гибкие и многофункциональные психометрические подходы, способные обеспечить создание инструмента, учитывающего разные типы понимания и временную динамику различных показателей понимания, которые, безусловно, изменяются под влиянием формального обучения и в процессе развития.

Ниже мы покажем, как использование «новой психометрики» (НП), основанной на моделях Раша, позволяет, в отличие от «классической психометрики» (КП), разрешить некоторые из описанных выше сложностей и создать тест на понимание прочитанного, который может использоваться при решении как практических, так и исследовательских задач.

Одно из принципиальных различий между НП и КП заключается в том, что в КП задания/утверждения/вопросы тестов и характеристики тех, кто отвечает на эти вопросы, неотделимы друг от друга. Иными словами, в КП индикатор утверждения/вопроса теста (так называемый показатель сложности или трудности утверждения) неотделим от индикатора тестируемого (так называемого уровня способностей тестируемого). В рамках КП разделение этих индикаторов невозможно, а вот в рамках НП модели Раша это разделение осуществимо. Важность такого разделения трудно переоценить [6; 18]: если два ученика

решают 10 задач нарастающей сложности и при этом один ученик решает 5 первых, более легких, задач, а второй – 5 последних, более трудных, то в КП их результаты будут одинаковыми (каждый из них получит по 5 баллов), а в НП их результаты будут отличаться (оценка латентной способности второго ученика будет выше, чем первого). Иными словами, в рамках НП оценка выполнения задания делается на основе двух параметров – параметра сложности определенного задания (вопроса или утверждения) и параметра уровня способности тестируемого.

НП по сравнению с КП характеризуется еще целым рядом преимуществ. Так, НП позволяет использовать альтернативные формы одного и того же инструмента или даже неперекрывающиеся наборы утверждений для получения гомогенной оценки латентной способности в заданной выборке [16]. Кроме того, НП позволяет отслеживать и при необходимости корректировать [17] равномерность использования шкалы Лайкерта, которая искажается либо отвечающими на вопросы, демонстрирующими предпочтение определенных позиций на шкале [10], либо при переходе от одного вопроса к другому. Наконец, НП предполагает использование более адекватных статистических разработок, когда сбор данных осуществляется гнездовым методом (например, когда данные учащихся «вложены» в данные их классного руководителя) и не все наблюдения являются независимыми друг от друга.

Примером таких «вложенных» данных является ситуация, когда при применении в тестировании открытых заданий (т.е. заданий, где тестируемый должен сам дать ответ, а не выбрать его из имеющегося набора ответов) привлекались оценщики (или рэйтеры). Для подобного рода данных используется так называемая Many-Facet Rasch Model (MFRM) [13]. Модель MFRM особенно часто применяется в тех случаях, когда одно и то же задание оценивается больше чем одним рэйтером [9] как для отдельно взятого тестируемого, так и для групп (или гнезд) тестируемых. Особое преимущество этой модели заключается в том, что она позволяет оценивать эффекты рэйтеров и учитывать эти эффекты при подсчете оценок латентной способности.

Мы использовали модель MFRM при оценивании процесса понимания. В проведенном нами исследовании приняли участие 4020 детей и взрослых из регионов России, где доминируют этнические славяне и русский язык. Выборка формировалась путем включения учащихся II–X классов (~40% девочки) и родителей примерно 25% детей. Возрастной диапазон выборки составил 7–68 лет.

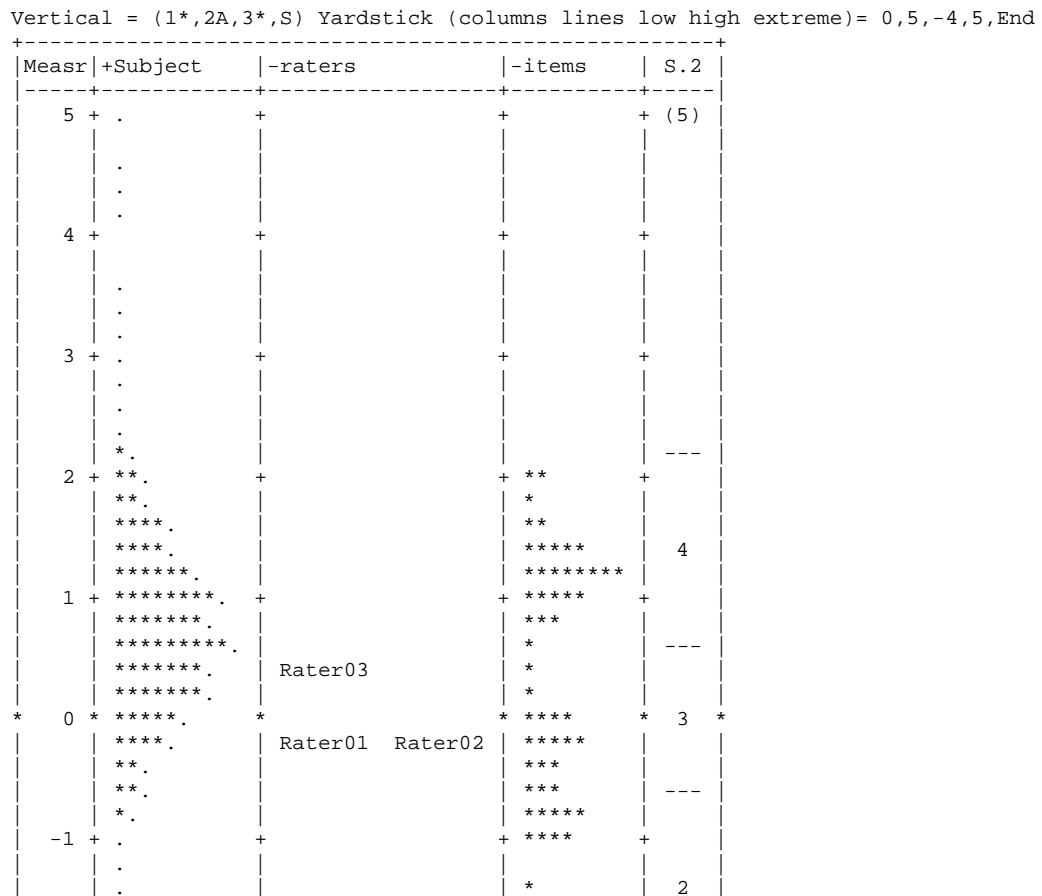
В качестве теста на понимание прочитанного использовался набор параграфов. Понимание параграфов оценивалось путем анализа ответов на вопросы в форматах множественного выбора и открытых заданий (от 1 до 6 вопросов/заданий для параграфа; всего 64 задания: 51 – на множественный выбор и 13 – открытых; задания были направлены как на фактическое, так и на инференциональное понимание). Всего было использовано 15 параграфов, которые были разбиты на четыре группы: 1) для учащихся II и III классов; 2) для учащихся IV–VI классов; 3) для учащихся VI–VIII классов; 4) для старшеклассников и взрослых. В каждом случае набор параграфов включал в себя пять текстов, уровни трудности которых внимательно контролировались согласно программе обучения русскому языку (<http://www.edusite.ru/p135aa1.html>). Кроме того, два параграфа из каждого набора параграфов предъявлялись не только той возрастной группе, для которой они разрабатывались, но и смежной возрастной группе, формируя «перекрытие», необходимое для применения MFRM (см. ниже). Иными словами, каждой возрастной группе предъявлялось только пять параграфов, но за счет перекрытия все возрастные группы были соединены в одной матрице данных.

Открытые задания обрабатывались по специально разработанным рубрикам, используя 5-балльную оценочную шкалу. В обработке принимали участие три группы рэйтеров: 1) пенсионеры, выполняющие работу по оцениванию по найму на основе почасовой нагрузки; 2) работающие люди среднего возраста, выполняющие работу по оцениванию по найму на основе почасовой нагрузки; 3) студенты и аспиранты, изучающие психологию. Все рэйтеры были специально подготовлены для процесса оценивания открытых ответов по рубрикам. В момент подготовки критерием их включения в последующую работу по оцениванию служило то, что согласованность их оценок с оценками тренера и, по крайней мере, еще с одним рэйтером достигала 70%. В последующем анализе данные рэйтеров внутри каждой группы были объединены в три суммирующих показателя, для того чтобы можно было отслеживать различия оценок, свойственные каждой из этих групп.

Согласно стандартной модели Раша, вероятность правильного ответа на определенное задание/утверждение/вопрос теста определяется взаимодействием двух факторов (фасеток) – уровнем способности тестируемого и уровнем трудности данного утверждения/вопроса теста [12; 18]. MFRM дальше развивает эту модель, вводя дополнительный фактор (фасетку) – оценку рэйтером ответа на открытую задачу [9]. Для

проверки этой последней модели, включающей три фактора – уровень способности тестируемого, уровень сложности задачи, фактор рэйтера, и параметризации этих факторов использовался пакет FACETS [8].

Как практически любой статистический пакет, позволяющий создавать сложные статистические модели, FACETS оценивает степень соответствия построенной модели тем эмпирическим данным, для которых эта модель предназначена. Так, согласно FACETS, моделью понимания прочитанного, построенной для описанных здесь данных, объясняется 76,6% дисперсии. На рис. 1 представлены распределение латентной способности понимания прочитанного в данной выборке, обобщенные позиции групп рэйтеров и распределение заданий и утверждений. Кроме того, пакет FACETS генерирует несколько статистических показателей для утверждений/заданий, показателей латентной способности тестируемых и характеристик оценок рэйтеров. Ниже мы последовательно проанализируем все эти показатели.



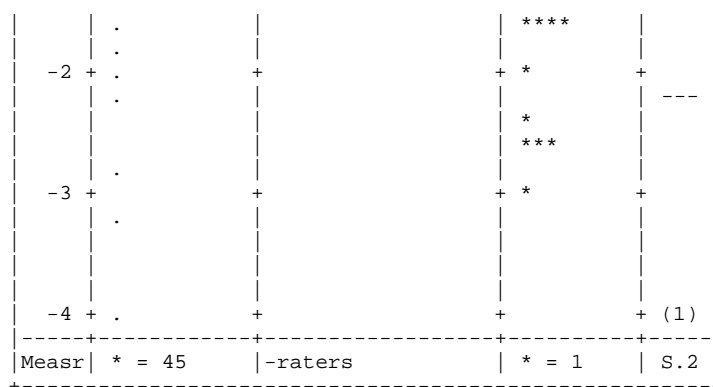


Рис. 1. Распределения тестируемых (левая колонка), рэйтеров (средняя колонка) и заданий/утверждений (правая колонка) на шкале латентной способности (по FACETS)

**Задания.** Показатели статистического соответствия (фита – fit) основаны на разнице между наблюдаемыми и ожидаемыми ответами, полученными для каждого тестируемого по каждому из заданий, вопросов или утверждений [6]. Локальные показатели соответствия могут рассматриваться как индикаторы наличия и величины «шума» в модели измерения. Для каждого задания оцениваются четыре показателя фита – средние квадратичные и резидуальные (или остаточные) величины для внутреннего (infit) и внешнего (outfit) индикаторов измерительной модели. Ожидаемые значения для средних квадратичных равны 1,0. Значения меньше 1,0 показывают тенденцию к повторяемости утверждений/заданий и избыточность предикторной информации; значения больше 1 показывают тенденцию к генерации немоделируемого «шума», уникальной изменчивости, которую невозможно объяснить, используя параметры модели. Однако само наличие этих тенденций не является критическим до того момента, пока средние квадратичные не достигнут определенных порогов. Величины в диапазоне от 0,50 до 1,50 считаются приемлемыми [7]. Если же средние квадратичные меньше 0,5 или больше 1,5, то соответствующие им задания должны быть проанализированы с намерением их изменить, заменить или удалить. Резидуальные (остаточные) величины являются дополнительным показателем степени отклонения полученных экспериментальных данных от ожидаемых модельных. Если средние остаточные величины не выходят за указанные выше пределы, то анализ остаточных величин проводить не обязательно.

Если же анализ резидуальных величин все-таки проводится, то маленькие остаточные величины свидетельствуют о хорошем соответствии между ожидаемым и наблюдаемым, а большие – о том, что полученный ответ оказался непредсказуемым, т. е. его нельзя было предсказать на основе теоретических предположений о том, как данный тестируемый, судя по его ответам на другие задания, должен был бы ответить на данное задание, учитывая уровень его сложности. При анализе остаточных величин принято пользоваться порогами  $-2$  и  $+2$  [14; 15]. Здесь, как и при анализе средних квадратичных величин, *infit* говорят об определенной степени редантности (повторяемости), а *outfit* – о присутствии «шума» в модели. Анализ показателей локальных индексов соответствия подтвердил адекватность проведения всех 64 заданий: все средние квадратичные для индексов *infit* оставались в диапазоне от 0,60 до 1,36, а для индексов *outfit* – от 0,47 до 1,38. Этот результат соответствует тому, что оценки надежности данной шкалы понимания составили 0,98. Сложности задания варьировались от 2,08 (самое сложное – индеференциальное задание на множественный выбор) до 2,96 (самое простое – фактические вопрос на множественный выбор).

**Рэйтеры.** Как указывалось выше, все рэйтеры, принимавшие участие в этой работе, были разделены на три группы, и для каждой из групп были получены оценочные показатели. С точки зрения данного исследования, такое объединение вполне допустимо, поскольку нужно было выяснить, во-первых, сопоставимость дисперсии фасетки тестируемых и фасетки рэйтеров (первая должны быть выше второй), во-вторых, степень сходства и рассогласования трех групп рэйтеров и учет этих особенностей при оценивании латентного фактора способностей (индивидуальные особенности каждого из рэйтеров интереса не представляли, соответственно, результаты MFRM показывают степень согласованности каждой из групп, принимавшей участие в оценивании).

Прежде всего отметим, как рэйтерами использовалась 5-балльная шкала. Проценты выбора каждого из баллов составили 15 (1), 25 (2), 38 (3), 16 (4) и 6 (5); оценка «5» использовалась меньше всего, но достаточно большое количество тестируемых все-таки получили этот самый высокий балл. На рис. 2 представлена зависимость оценок рэйтеров от уровня латентной способности. Далее, обратим внимание на то, что дисперсия среди тестируемых была намного больше, чем дисперсия среди рэйтеров ( $0,74 \pm 0,90$  и  $0,00 \pm 0,34$ , соответственно). Причем все три группы рэйтеров показали относительно высокую степень согласованности – 56,1% по сравнению с ожидаемым уровнем согласованности – 32,3%; это

говорит о том, что рэйтеры достаточно систематично оценивают открытые вопросы (наблюдаемый показатель примерно в два раза превышает показатель ожидаемый). И, наконец, выявилось отличие между тремя группами рэйтеров ( $\chi^2(2) = 1194,7$ ;  $p < 0,01$ ). Из трех групп рэйтеров две – пенсионеры и работающие люди – оказались более похожими друг на друга, в то время как группа студентов и аспирантов, специализирующихся в области психологии, значительно отличалась от этих групп. Студенты и аспиранты оказались самыми критическими «ценителями» (логит 0,39), в то время как две остальные группы были близки друг к другу (логиты 0,14 и 0,25 для пенсионеров и работающих людей, соответственно).

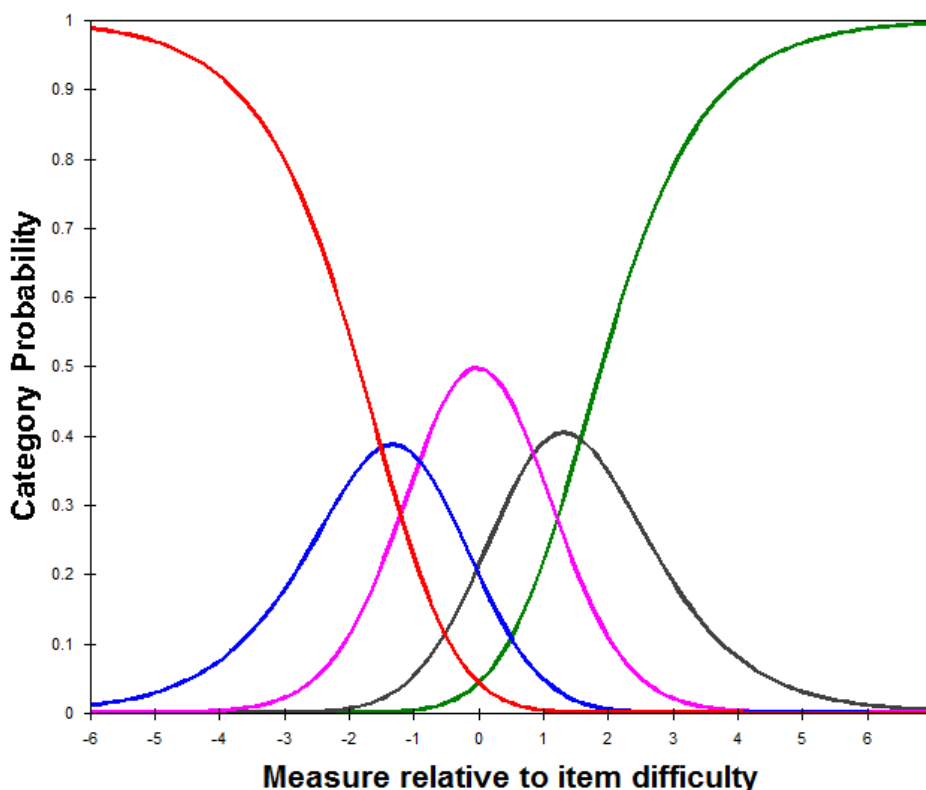


Рис. 2. Зависимость оценок рэйтеров от уровня латентной способности (по FACETS)

**Тестируемые.** Главным преимуществом модели MFRM является ее способность разделять характеристики утверждений/вопросов тестов и характеристики тех, кто отвечает на эти вопросы. Характеристики тестируемых моделируются отдельно на шкале латентной способности (см. рис. 1).



Как и для остальных фасеток, для фасетки тестируемых FACETS приводит несколько статистик, свидетельствующих о характеристиках инструмента. Основной показатель здесь – показатель надежности теста. Этот показатель подсчитывается два раза – для всего теста и затем для теста, из которого удалены так называемые «экстремальные» задания/утверждения, выполненные определенными тестируемыми (так называемые экстримы). В данном случае первая величина составила 0,68, а вторая – 0,71. Очевидно, что удаление нескольких экстримов не влияет на показания по шкале; полученная шкала надежна.

Еще одним информативным показателем является показатель «страты» (strata), который говорит о том, сколько групп тестируемых данный тест различает статистически. В данном случае страта для всего теста – 2,28, а страта для тестов с удаленными экстримами – 2,43. Эти показатели похожи друг на друга. Так, несмотря на то, что в данной выборке по тесту на понимание прочитанного существуют экстримы, они несущественно влияют на показатели теста. Величины страт говорят о том, что тест статистически надежно различает слабых, средних и сильных тестируемых. Диапазон латентной способности испытуемых в данной выборке варьировался от 4,14 (эта самая высокая оценка была дана учащейся II класса) до 5,36 (эта самая низкая оценка была дана учащемуся II класса).

Таким образом, использованный нами тест на понимание прочитанного, будучи проанализированным с помощью MFRM, обладает адекватными психометрическими свойствами, покрывает большой диапазон возрастов при возможности проведения тестирования в относительно короткий срок разными перекрывающимися формами теста и позволяет статистически дифференцировать, по крайней мере, три группы тестируемых. Однако возникает вопрос: как сопоставляются результаты по этому тесту, полученные в контексте НП (MFRM), с результатами, полученными в контексте КП (т.е. методом простого подсчета правильных и неправильных ответов)?

Для получения ответа на этот вопрос было проведено два анализа.

Во-первых, был подсчитан простой суммарный балл по тесту на понимание. Этот балл был прокоррелирован с оценкой латентной способности, полученной методом MFRM. Корреляция составила  $r = 0,32$  ( $p < 0,001$ ). Очевидно, что два метода переработки информации, полученной от тестируемых, даже когда данные собраны на одном и том же тестовом материале и в одной и той же выборке, дают суммарные баллы, которые весьма далеки друг от друга. Без сомнения, это – важная информация к размышлению, говорящая о

том, что простой подсчет правильных ответов неадекватно отражает уровень процесса понимания тестируемых.

Во-вторых, для того, чтобы продемонстрировать преимущество показателей, полученных методами НП (MFRM), по сравнению с показателями, полученными методами КП, из большой выборки была выбрана небольшая группа испытуемых, состоящая из нескольких подгрупп по классам, в которых они обучаются. При анализе методами КП в силу отличий между выполняемыми заданиями и несмотря на то, что для каждого набора существовали перекрывающиеся номера, данные, полученные в разных классах школы (в IV, VI и VIII классах), сопоставимы и контр-интуитивны, поскольку средние для каждой из групп составили: для IV класса – 74,74; для VI класса – 44,26; для VIII класса – 85,55. Средние на латентной шкале способностей, однако, сравнимы друг с другом и полностью соответствуют ожиданиям: среднее для IV класса равно 1,13; среднее для VI класса равно 1,45; среднее для VIII класса равно 1,54.

Как уже говорилось выше, КП не показывает этиологии заданий/утверждений: один и тот же показатель по КП может занимать две разные позиции на шкале латентной способности, произведенной в контексте НП. MFRM предлагает несколько вариантов интерпретации ко-распределения баллов, полученных для заданий/утверждений и для латентной шкалы способностей. Одним из таких вариантов является анализ «неожиданных ответов» – ответов конкретных испытуемых на конкретное задание в ключе, который не соответствует паттерну ответов, зарегистрированных для этого тестируемого.

Итак, мы коротко описали основные принципы MFRM и проиллюстрировали эти принципы. Применение MFRM при работе с инструментарием для изучения процесса понимания прочитанного позволяет сделать следующие выводы. Во-первых, стало ясно, как создать первичную базу данных, с помощью которых можно исследовать процесс понимания прочитанного более-менее систематически, опираясь на набор параграфов из данного исследования как на некий базовый, и постоянно дополнять банк параграфов (и, соответственно, заданий) новыми единицами или заменять старые неудачные, с психометрической точки зрения, параграфы и задания новыми. Во-вторых, выяснилось, что НП, в отличие от КП, позволяет объединить в одном анализе ответы на вопросы множественного выбора и открытые утверждения. В-третьих, стало возможным применять одну шкалу латентных способностей при одновременном тестировании людей очень разных

возрастов. Это огромное преимущество НП по сравнению с СП при исследовании выборок, включающих несколько поколений (например, выборок, состоящих из семей).

Исследования, в контексте которых были собраны эмпирические материалы, приведенные в данной статье, финансово поддерживались следующими спонсорами: CRDF, NIH (DC007665 and HD052120), International Dyslexia Association. Автор выражает благодарность всем своим коллегам, которые помогли в сборе эмпирического материала, в предварительной подготовке этого материала и в проведении MFRM-анализа. Особая благодарность – всем участникам исследования за выделенное ими время и вложенные усилия.

## Литература

1. *Болотов В., Ковалева Г.* Опыт России в области оценки образовательных достижений школьников [Online]. 20 августа 2011 г.  
<http://ciced.ru/docs/publications/V.%20Bolotov.%20Opit%20Rossii%20v%20oblasti%20ocenki%20obrazovatelnih%20dostizheniy%20shkolnikov.pdf>.
2. *Гончарова Е.Л.* Ранние этапы становления читательской деятельности в норме и при отклонениях в развитии. М., 2009.
3. *Корнев А.Н.* Нарушения чтения и письма у детей. СПб., 2003.
4. Постановление ЦК ВКП(б) от 04 июля 1936 г. «О педологических извращениях в системе наркомпросов». М, 1936.
5. *Цукерман Г.А., Ермакова И.В., Кудина Г.Н., Соколова О.В.* Понимание противоречий: микроанализ задач теста PISA-2000 // Психологическая наука и образование. 2005. №1.
6. *Bond T.G., Fox C.M.* Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahwah, NJ, 2001.
7. *Linacre J.M.* Facets Rasch Measurement Computer Program Manual (Version 3.65.0). Chicago, IL., 2009.
8. *Linacre J.M.* Facets Rasch Measurement Computer Program. Chicago, IL, 2004.
9. *Linacre J.M., Wright B.D.* Construction of Measures from Many-facet Data // Journal of Applied Measurement. 2002. Vol. 3.

10. *Lopez W.* Communication Validity and Rating Scales // Rasch Measurement Transactions. 1996. Vol. 10.
11. *Oakhill J., Cain K.* Introduction to Comprehension Development // Children's Comprehension Problems in Oral and Written Language: A Cognitive Perspective / ed. K. Cain, J. Oakhill. N. Y., NY, 2007.
12. *Rasch G.* An Item Analysis that Takes Individual Differences into Account // The British Journal of Mathematical and Statistical Psychology. 1966. Vol. 19.
13. *Smith E.V.J., Conrad K.M., Chang K., Piazza J.* An Introduction to Rasch Measurement for Scale Development and Person Assessment // Journal of Nursing Measurement. 2002. Vol. 10.
14. *Smith R.M.* Fit Analysis in Latent Trait Measurement Models // Journal of Applied Measurement. 2000. Vol. 1.
15. *Smith R.M., Schumacker R.E., Bush J.J.* Examining Replication Effects in Rasch Fit Statistics // Objective Measurement: Theory into Practice / ed. M. Wilson, G. Engelhard. Stamford, CT, 2000. Vol. 5.
16. *Wolfe E.W.* Equating and Item Banking with the Rasch Model // Journal of Applied Measurement. 2000. Vol. 1.
17. *Wright B.D.* Logits? // Rasch Measurement Transactions. 1993. Vol. 7.
18. *Wright B.D., Mok M.* Rasch Models Overview // Journal of Applied Measurement. 2000. Vol. 6.

---

# New psychometrics and evaluation of understanding of the read material

*E. L. Grigorenko,*

*PhD in psychology, professor, head of laboratory for behavior genetics, chair of educational psychology and pedagogics, department of psychology, Moscow State University named after M.V.Lomonosov elenalgrigorenko@gmail.com*

---

This article (a brief report) discusses the main characteristics of the Many-Facet Rasch Model, (MFRM) and illustrates the use of this model by the example of a research of individual differences in the process of understanding of the read material. The illustration is based on the data from 4020 testees whose responses to the questions on the read material were integrated in a single latent skill scale, despite a vast range of age-specific characteristics of the testees and the variety of questions\tasks..

**Keywords:** process of understanding of the read material, classic and new psychometrics, Rasch Model fit indices, Many-facet models

---

## References

1. *Bond T. G., Fox C. M.* Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah, NJ: Erlbaum, 2001.
2. *Linacre J. M.* Facets Rasch measurement computer program. Chicago, IL: Winsteps.com, 2004.
3. *Linacre J. M.* Facets Rasch measurement computer program manual (version 3.65.0). Chicago, IL: Winsteps.com, 2009.
4. *Linacre J. M., Wright B. D.* Construction of measures from many-facet data // Journal of Applied Measurement. 2002. Vol. 3. P. 486-512.
5. *Lopez W.* Communication validity and rating scales // Rasch Measurement Transactions. 1996. Vol. 10. P. 482.

6. *Oakhill J., Cain K.* Introduction to comprehension development // Children's comprehension problems in oral and written language: A cognitive perspective / ed. K. Cain, J. Oakhill. New York, NY: Guilford Press, 2007. P. 3-40.
7. *Rasch G.* An item analysis that takes individual differences into account // The British Journal of Mathematical and Statistical Psychology. 1966. Vol. 19. P. 49-57.
8. *Smith E. V. J., Conrad K. M., Chang K., Piazza J.* An introduction to Rasch measurement for scale development and person assessment // Journal of Nursing Measurement. 2002. Vol. 10 P. 189-206.
9. *Smith R. M.* Fit analysis in latent trait measurement models // Journal of Applied Measurement. 2000. Vol. 1. P. 199-218.
10. *Smith R. M., Schumacker R. E., Bush J. J.* Examining replication effects in Rasch fit statistics // Objective measurement: Theory into practice / ed. M. Wilson, G. Engelhard. Stamford, CT: Ablex Publishing Corp, 2000: Vol. 5. P. 303-317.
11. *Wolfe E. W.* Equating and item banking with the Rasch model // Journal of Applied Measurement. 2000. Vol. 1. P. 409-434.
12. *Wright B. D.* Logits? // Rasch Measurement Transactions. 1993. Vol. 7. P. 288.
13. *Wright B. D., Mok M.* Rasch models overview // Journal of Applied Measurement. 2000. Vol. 6. P. 83-106.
14. *Bolotov V., Kovaleva G.* Opyt Rossii v oblasti ocenki obrazovatel'nyh dostizhenii shkol'nikov [Online]. 20 avgusta, 2011. <http://ciced.ru/docs/publications/V. Bolotov. Opit Rossii v oblasti ocenki obrazovatelnih dostizheniy shkolnikov.pdf>.
15. *Goncharova E. L.* Rannie etapy stanovleniya chitatel'skoi deyatelnosti v norme i pri otkloneniyah v razviti. Moskva: Institut Korrekcionnoi Pedagogiki, 2009.
16. *Kornev A. N.* Narusheniya chteniya i pis'ma u detei. Sankt-Peterburg: Rech', 2003.
17. *CK VKP(b)* Postanovlenie CK VKP(b) ot 04.07.1936 o pedologicheskikh izvrasheniyah v sisteme NARKOMPROSOV. Moskva, 1936.
18. *Cukerman G. A., Ermakova I. V., Kudina G. N., Sokolova O. V.* Ponimanie protivorechii: mikroanaliz zadach testa PISA-2000 // Psihologicheskaya nauka i obrazovanie 2005. Vol.