

---

**ИССЛЕДОВАНИЯ В ОРГАНИЗАЦИОННОЙ ПСИХОЛОГИИ**

---

## Прогностическая валидность Центра оценки: метаанализ<sup>1</sup>

**ХЕРМЕЛИН Иран**

*Психологическая группа, Manchester Business School, Университет Манчестера (Манчестер, Великобритания)*

*E-mail: info@eranhermelin.com*

**ЛАЙВЕНС Филип**

*Факультет управления персоналом и психологии труда и организационной психологии, Гентский университет (Гент, Бельгия)*

**РОБЕРТСОН Айвен**

*Robertson Cooper Ltd. (Манчестер, Великобритания)*

Перевод с англ. Е. Лурье (руководитель центра разработки оценочных технологий «ЭКОПСИ Консалтинг») и В. Тихомировой (студентка факультета психологии НИУ ВШЭ).

*Данное метааналитическое исследование посвящено оценке прогностической критериальной валидности Центров оценки (ЦО). Его цель – обновить ранее полученные исследовательские данные о валидности ЦО. Нами были проанализированы 26 исследований и 27 коэффициентов валидности ( $N = 5850$ ), связывающие общий оценочный рейтинг (ООР)<sup>2</sup> с рейтингом результативности (РР)<sup>3</sup>. Корреляция между ООР и РР после коррекции данных составила 0.28 (95%-ный доверительный интервал =  $0.24 \leq \rho \leq 0.32$ ). Кроме того, делается предположение, что такая оценка валидности не является окончательной, так как показатели валидности, как правило, зависят от косвенного ограничения дисперсии.*

---

1 Эта статья написана на базе тезисов, взятых из работы Ирана Хермелина. Работа была выдвинута для защиты степени PhD по менеджменту в Институте науки и технологии Манчестерского университета (University of Manchester Institute of Science and Technology).

**Источник:** *Hermelin E., Lievens F., Robertson I.T. The Validity of Assessment Centres for the Prediction of Supervisory Performance Ratings: A meta-analysis // International Journal of Selection and Assessment. 2007. Vol. 15. Is. 4. P. 405–411. Публикуется с разрешения издательства John Wiley and Sons.*

2 Общий оценочный рейтинг, ООР (от англ. Overall Assessment Rating, OAR) – итоговый рейтинг участников Центра оценки. Формируется на основе оценок обученных наблюдателей-экспертов (*здесь и далее прим. пер.*).

3 Рейтинг результативности, РР (от англ. Supervisory Performance Rating, SPR) – рейтинг сотрудников, составленный по результатам их деятельности в организации. Формируется на основании оценок руководителей сотрудников или их супервизоров.

## Введение

В управлении человеческими ресурсами Центр оценки используют для решения двух основных задач [Thornton, 1992]. Во-первых, традиционно он применяется для отбора и продвижения персонала. В этом случае при принятии решений главную роль играет общий оценочный рейтинг. В метаанализе, выполненном Гоглер, Розенталем, Торнтоном и Бентсоном [Gaugler, Rosental, Thornton, Bentson, 1987], приводятся данные о высокой валидности прогнозов, сделанных на основе ООР. В частности, среднее значение прогностической валидности ООР в их исследовании составило 0.36. Критериальной мерой для оценки валидности являлся рейтинг результативности.

Во-вторых, ЦО все чаще используется для развития управленческих навыков (ЦР – центр развития). В этом случае наиболее значимыми для принятия решений становятся оценки по отдельным компетенциям, ООР же отходит на второй план. Анализируя результаты по компетенциям, участники могут получить исчерпывающую информацию о своих сильных и слабых сторонах. В 2003 г. Артур, Дэй, МакНелли и Иденс [Arthur, Day, McNelly, Edens, 2003] опубликовали работу, результаты которой подтверждают критериальную валидность измерений отдельных компетенций в ЦО. В своем метаанализе исследователи проанализировали оценки по шести наиболее значимым компетенциям (уважение/понимание других, коммуникабельность, энергичность, влияние, организованность и планирование, решение проблем). Значения их прогностической валидности варьировались в диапазоне от 0.25 до 0.39.

Несмотря на то, что в 2003 г. было опубликовано исследование прогностической валидности измерения компетенций в ЦО [Arthur et al., 2003], по вопросу валидности общего оценочного рейтинга «золотым стандартом» до сих пор так и оставалось исследование Гоглер [Gaugler et al., 1987]. Но со времени этой публикации появились новые данные о валидности ООР, которые мы и включили в наше исследование. В данном метаанализе проанализированы работы, выполненные с 1985 по 2005 г. Как и в исследовании Гоглер, ЦО в данной работе рассматривался с точки зрения способности отбирать лучших кандидатов на имеющуюся должность. В качестве критериальной меры, отражающей эффективность ЦО, использовался рейтинг результативности. Валидность измерения отдельных компетенций не рассматривалась [Arthur et al., 2003].

## Методы

### Базы данных

Для поиска подходящих исследований мы использовали несколько стратегий. Во-первых, был проведен поиск по электронным библиографическим базам данных (PsycInfo, Social Sciences Citation Index и т.п.). Во-вторых, применялся поиск по электронной базе данных Британского психологического общества, в которой содержатся контакты аккредитованных британских психологов. От ученых и практиков были получены контакты специалистов, имевших доступ к неопубликованным данным о валидности ЦО. Кроме того, для получения необходимой информации авторы связались с 20 ведущими

компаниями из индекса FTSE-500<sup>4</sup> и с четырьмя крупнейшими британскими фирмами, оказывающими услуги в области организационной психологии.

### **Критерии включения**

В исследование мы включали только те данные, которые соответствовали следующим четырем критериям. Во-первых, исследование рассматривалось, если оно было посвящено именно ЦО. Под ЦО в данном случае понималось следующее:

- а) в процедуру оценки должно быть включено не менее двух методов. При этом как минимум один из этих методов должен быть индивидуальным или групповым симуляционным упражнением;
- б) как минимум в одном из симуляционных упражнений должен открыто присутствовать один или несколько наблюдателей-оценщиков;
- в) оценки действий участников сводятся в общий оценочный рейтинг клинически (экспертным путем) и/или статистически;
- г) минимальное время проведения процедуры оценки – 2 часа.

Так как мы стремились обновить предыдущие исследования, вторым требованием стало временное ограничение круга рассматриваемых работ. Во внимание принимались исследования, опубликованные или завершённые с 1985 по 2005 г.

В-третьих, в качестве критериальных мер мы использовали только те данные, которые соответствовали определению рейтинга результативности, предложенному Борманом [Borman, 1991]. Под РР понималась оценка результативности работника, полученная от его руководителя или супервизора. При этом результативность могла быть оценена как интегрально, так и дифференцированно по набору критериев. Оценка потенциала сотрудника, объективные показатели его результативности, результаты по тестам достижений, а также оценки, сделанные коллегами, не рассматривались как основания для формирования рейтинга.

В-четвертых, исследование должно было содержать достаточное количество информации об условиях сбора данных. В преобладающем большинстве случаев всей необходимой информации в работах не доставало. Для того чтобы ее уточнить, Иран Хермелин пытался связываться непосредственно с авторами исследований.

В результате в метаанализ были включены 26 независимых исследований с 27 коэффициентами валидности. Итоговая выборка составила 5850 человек (по сравнению с выборкой в 4180 в исследовании Гоглер [Gaugler et al., 1987]). В список вошло 23 опубликованных исследования, из них одно, представленное на международной конференции, и два неопубликованных. Самое раннее исследование датировалось 1985 г., самое позднее – 2005 г.

Кодированием 27 коэффициентов валидности, которые легли в основу итогового метаанализа, отдельно занимались два автора настоящей статьи – Хермелин и Лайвенс. Проверка надежности при сопоставлении результатов работ авторов показала совпадение кодирования на 85%. За основу финальной версии была взята схема кодирования, предложенная первым автором. В завершении процедуры авторы сравнили независимо полученные результаты и приняли согласованное решение по каждому из расхождений.

4 FTSE-500 Index (англ. Financial Times Stock Exchange Index) — британский фондовый индекс. Рассчитывается агентством FTSE Group на основе курсов акций компаний с наибольшей капитализацией, включенных в список Лондонской фондовой биржи (LSE). Считается одним из наиболее влиятельных биржевых индикаторов в Европе.

### **Оценка степени влияния систематической ошибки**

Мы проверили полученную базу данных дополнительно по двум параметрам. Первый параметр – наличие или отсутствие неучтенных побочных переменных, таких как время. Для этого исправленные показатели валидности были соотнесены с датой публикации/готовности исследования, а также с временным интервалом между составлением ООР и РР. Показатели корреляции составили  $r = -0.05$  и  $0.10$  соответственно. Таким образом, показатели валидности ЦО не были значимо связаны со временем.

Вторым проверяемым параметром являлось влияние систематической ошибки на выборку. Для проверки распределения скорректированных показателей валидностей была построена воронкообразная диаграмма рассеяния (см. Egger et al., 1997). Эта процедура состоит в построении графика, по оси абсцисс которого указывается объем выборки, а по оси ординат – исправленные валидности. Используя такой график, можно увидеть, зависят ли показатели валидности от систематической ошибки. Так как величина ошибки выборки зависит от ее размера, разброс коэффициентов валидности должен ожидаемо уменьшаться по мере увеличения объема выборки. В связи с этим диаграмма рассеяния будет иметь вид симметричной перевернутой воронки. Если же имеет место ошибка выборки, то воронкообразный график будет асимметричным [Egger et al., 1997]. Диаграмма рассеяния для коэффициентов валидности, полученных на небольших выборках, напоминала перевернутую воронку. Разброс значений здесь оказался высоким, тогда как коэффициенты, полученные на больших выборках, группировались ближе к среднему значению валидности. При этом данные все же оказались распределены неравномерно относительно среднего. Значения шести коэффициентов располагались ниже полученного в метаанализе среднего значения, один коэффициент оказался равен среднему и 20 коэффициентов располагались выше среднего. В исследованиях с большим объемом выборки коэффициенты валидности обычно располагались кучнее вокруг метааналитического среднего по сравнению с работами с малым объемом выборки. Исследование с наибольшим объемом выборки (28% общего объема) располагалось в середине распределения и поэтому не искажало итоговый результат метаанализа.

### **Процедура метаанализа**

Полученная база данных анализировалась при помощи процедуры, предложенной Хантером и Шмидтом [Hunter, Schmidt, 1990]. Суть ее заключалась в индивидуальном исправлении корреляций с учетом выявленных экспериментальных артефактов. В случаях когда данных, полученных в ходе анализа, оказывалось недостаточно, мы обращались за необходимой информацией к книге второго автора [Schmidt, личное общение, 2002]. В ходе анализа для 20 из 27 коэффициентов валидности мы смогли выявить факторы, ограничивающие дисперсию ООР. Затем каждый из 20 коэффициентов был закодирован с учетом этих ограничений. По причине неполноты информации для оставшихся 7 коэффициентов это значение было задано искусственно – путем вычисления среднего ограничения дисперсии по предыдущим 20 коэффициентам (см. приложение А).

Так как надежность рейтинга результативности во включенных в метаанализ исследованиях, как правило, обходилась стороной, мы решили использовать лучшие из уже имеющихся данных оценки надежности РР. В двух масштабных метааналитических исследованиях по этой проблеме было получено среднее значение надежности РР, равное 0.52

[Salgado et al., 2003; Viswesvaran et al., 1996]. Таким образом, мы решили использовать данное значение в качестве коэффициента надежности РР для всех 27 коэффициентов валидности.

В настоящее время существуют методики, с помощью которых можно корректировать данные при неявно ограниченной дисперсии [Hunter et al., 2006; Schmidt et al., 2006]. Однако в большинстве проанализированных нами исследований отсутствовала информация о параметрах, косвенно ограничивающих дисперсию. По этой причине мы не имели права отойти от стандартной процедуры «очистки» значений валидностей от ненадежности критерия и прямого ограничения дисперсии.

## Результаты

### Валидность ЦО

По данным, представленным в табл. 1, видно, что среднее значение коэффициента корреляции по итоговой выборке ( $N = 5850$ ) составило 0.17. «Очистка» этого коэффициента от факторов, прямо ограничивающих дисперсию ООР, привела к увеличению показателя до 0.20. После того как была проведена коррекция на ненадежность критерия (РР), значение корреляции между ООР и РР в целом по выборке повысилось до 0.28 (95%-ный доверительный интервал =  $0.24 \leq \rho \leq 0.32$ ). Данные по артефактам, приводившим к систематической ошибке, представлены в табл. 1. Эти данные использовались для корректировки коэффициентов валидности. По ним видно, что 84% дисперсии коэффициентов могут объясняться ошибкой выборки. Следовательно, даже после устранения влияния

**Таблица 1.** Краткое изложение результатов мета-анализа

<i>Средние результаты по валидностям мета-аналитической базы данных</i>								
<i>N</i>	<i>K</i>	<i>r</i>	<i>r<sub>c</sub></i>	<i>r<sub>cb</sub>(ρ)</i>	<i>95% ДИ</i>			
5850	27	.17	.20	.28	.24 ≤ ρ ≤ .32			
<i>Распределение мета-аналитических артефактов, используемых для корректировки исходных корреляций</i>								
<i>U</i>	<i>σ<sub>u</sub><sup>2</sup></i>	<i>c</i>	<i>σ<sub>c</sub><sup>2</sup></i>	<i>r<sub>yy</sub></i>	<i>Σ<sub>ryy</sub><sup>2</sup></i>	<i>b</i>	<i>σ<sub>b</sub><sup>2</sup></i>	<i>Ā</i>
.94	.016	.94	.015	.52	0	.72	0	.64
<i>Дисперсия значений для мета-аналитической базы данных</i>								
<i>σ<sub>rbc</sub><sup>2</sup></i>	<i>σ<sub>e_rbc</sub><sup>2</sup></i>	<i>σ<sub>ρ</sub><sup>2</sup></i>	<i>Объясняемая дисперсия</i>			<i>SD<sub>ρ</sub></i>		
.0123	.0104	.019	84%			.04		

**Примечания:** *N* – величина выборки в данном мета-анализе; *K* – количество коэффициентов валидности в данной выборке; *r* – среднее взвешенное наблюдаемой корреляции; *r<sub>c</sub>* – среднее взвешенное наблюдаемой корреляции с учетом диапазона ограничений; *r<sub>cb</sub>(ρ)* – среднее взвешенное наблюдаемой корреляции с учетом ограниченной дисперсии и ненадежности критерия; 95% ДИ – 95%-ный доверительный интервал; *U* – среднее значение ограниченной дисперсии в ООР; *σ<sub>u</sub><sup>2</sup>* – разброс значений ограниченной дисперсии в ООР; *c* – среднее значение фактора коррекции ограниченной дисперсии; *σ<sub>c</sub><sup>2</sup>* – разброс значений фактора коррекции ограниченной дисперсии; *r<sub>yy</sub>* – среднее значение критерия надежности; *Σ<sub>ryy</sub><sup>2</sup>* – разброс значений критерия надежности; *b* – среднее значение фактора коррекции; *σ<sub>b</sub><sup>2</sup>* – разброс значений фактора коррекции ненадежности критерия; *Ā* – среднее значение фактора затухания (*b* \* *c*); *σ<sub>rbc</sub><sup>2</sup>* – дисперсия валидностей, скорректированных с учетом диапазона ограничения и ненадежности критерия; *σ<sub>e\_rbc</sub><sup>2</sup>* – взвешенное среднее ошибки выборки, оцененное с учетом валидностей, скорректированных для диапазона ограничения и ненадежности критерия; *σ<sub>ρ</sub><sup>2</sup>* – остаточная дисперсия валидностей, скорректированных с учетом диапазона ограничений и ненадежности критерия; Объясняемая дисперсия – процент дисперсии, объясненный ошибкой выборки; *SD<sub>ρ</sub>* – стандартное отклонение скорректированных корреляций.

ошибки выборки часть необъясненной дисперсии все равно оставалась. Поиск побочных переменных, влияющих на данные, не привел к результату. Тем не менее были учтены такие переменные, как количество оцениваемых компетенций, количество оригинальных отборочных методик и тип процедуры, по которой происходило включение. Как и следовало ожидать, влияния не оказывала ни одна из перечисленных переменных.

## Обсуждение результатов

Традиционная процедура оценки валидности общих оценочных рейтингов, получаемых в результате Центров оценки, построена на базе исследований, проведенных до 1985 г. Некоторым работам, включенным в предыдущие исследования, около 50 лет. Однако за последние 20 лет было проведено большое количество новых исследований. Помимо обновления сведений о валидности отдельных компетенций ЦО, сделанного Артуром [Arthur et al., 2003], необходимо также провести обновление прогностической валидности общих оценочных рейтингов. Это важно, так как ООР чаще всего используется в ЦО (ООР в ЦР не играет такой значимой роли). Проведенный метаанализ призван дополнить информацию, полученную Гоглер [Gaugler et al., 1987]. Кроме того, настоящее исследование выполнено с привлечением более объемной выборки ( $N = 5850$ ), по сравнению с  $N = 4180$ , представленной в метаанализе Гоглер [Gaugler et al., 1987].

Среднее значение корреляции между ООР и РР составило  $\rho = 0.28$  (95%-ный доверительный интервал =  $0.24 \leq \rho \leq 0.32$ ). Таким образом, коэффициент корреляции в данном исследовании оказался значимо ниже, чем коэффициент, представленный в исследовании Гоглер [Gaugler et al., 1987], где  $\rho = 0.36$  (95%-ный доверительный интервал =  $0.30 \leq \rho \leq 0.42$ ). Ее коэффициент лежит вне установленного нами 95%-ного доверительного интервала. Возможным объяснением такой разницы может стать тот факт, что в современных ЦО участники проходят более тщательный отбор (поэтому, кстати, ЦО такие дорогие), чем раньше. А это приводит к большему косвенному ограничению дисперсии и, следовательно, к более низким значениям валидности до и даже после коррекции.

К сожалению, «очистка» данных от косвенного ограничения дисперсии невозможна по причине отсутствия подобного рода информации в рассмотренных исследованиях. Тем не менее при дополнительном анализе были найдены некоторые неявные указания на влияние косвенного ограничения диапазона оценок участников ЦО. В частности, в 6 исследованиях, включенных в нашу матрицу данных, имелась информация о валидности тестов способностей. При этом тесты использовались в процедурах отбора наряду с ЦО или же в самом ЦО. Средний показатель валидности тестов способностей составил 0.10 ( $N = 1757$ ) (в качестве критерия использовался РР). По данным, полученным в США и описанным Хантером, Шмиттом, Гудингом, Нои и Кирчем [Hunter, 1983; Schmitt, Gooding, Noe, Kirsch, 1984], валидность тестов способностей используемых в ЦО или наряду с ЦО, значительно ниже, чем валидность тестов способностей, используемых в качестве самостоятельных предикторов (0.22 и 0.24). По данным современных европейских исследований [Salgado et al., 2003], полученное в нашем исследовании значение валидности также значительно ниже валидности тестов способностей, применяемых независимо (0.29). Хотя проводить параллели следует с осторожностью, все же снижение уровня валидности тестов способностей, используемых в ЦО или наряду с ЦО, может быть следствием предварительного

отбора по когнитивным показателям. А это является значительным косвенным ограничением дисперсии ООР.

Исходя из результатов исследования в целом, можно сделать следующий вывод: при анализе валидности критериев необходимо принимать во внимание тот факт, что отбор персонала проходит в несколько этапов [Hermelin, Robertson, 2001; Roth et al., 2001]. В связи с этим мы рекомендуем будущим исследователям ЦО учитывать и скрупулезно прописывать следующие условия проведения исследований:

- 1) этап отбора персонала, на котором использовался ЦО;
- 2) результаты оценки участников на предварительных этапах отбора;
- 3) корреляцию между прогнозом, сделанным на предварительном отборочном этапе, и между ООР, используемым на более поздних этапах.

Так что более полный анализ косвенного ограничения дисперсии в ЦО станет возможным только после того, как будут доступны перечисленные данные. Также это позволит скорректировать имеющиеся данные с учетом косвенного ограничения дисперсии (в соответствии с процедурой, предложенной Хантером и Шмидтом [Hunter et al., 2006; Schmidt et al., 2006]).

Что касается влияния побочных переменных, то их контроль обеспечен низким процентом необъясненной дисперсии после коррекции на ошибку выборки. Таким образом, представление о том, что ЦО должен показывать значительный разброс значений коэффициентов валидности из-за разности в подходах к разработке и проведению, оказалось несостоятельным. Мы полагаем, что подобные результаты обусловлены не отсутствием влияния побочных переменных, а низкой вероятностью разброса значений в имеющихся данных.

Внимания будущих исследований прогностической валидности ЦО заслуживают следующие проблемы. Во-первых, это необходимость расширения и доработки критериев, по которым происходит измерение валидности ЦО. Например, одна из ведущих разработок в критериальной теории за последнее десятилетие посвящена выявлению взаимосвязей между эффективностью деятельности и организационной гражданственностью [Borman, Motowidlo, 1993]. Насколько нам известно, исследований на тему взаимосвязи результатов ЦО с поведением человека как гражданина не проводилось. Это странно, поскольку одним из ключевых преимуществ ЦО является его способность измерять межличностные компетенции.

Во-вторых, крайне важно исследовать конкурентную валидность ЦО по сравнению с симуляционными упражнениями с низкой экологической валидностью, такими как кейс-тесты [McDaniel et al., 2001]. Эти инструменты стали популярны по причине невысокой стоимости и легкости проведения оценки для большого количества участников. Кроме того, они вполне применимы для оценки межличностных критериев и демонстрируют хорошую прогностическую валидность.

### **Благодарности**

Мы выражаем благодарность Фридриху Ансилу (Frederik Anseel) за его проницательные комментарии к черновой версии статьи.

## Литература

- Arthur W. Jr., Day E.A., McNelly T.L., Edens P.S.* A Meta-Analysis of the Criterion-Related Validity of Assessment Center Dimensions // *Personnel Psychology*. 2003. Vol. 56. P. 125–154.
- Borman W.C.* Job Behavior, Performance, and Effectiveness // *Dunnette M.D., Hough L.M.* (eds) *Handbook of Industrial and Organizational Psychology*. Palo Alto, CA: Consulting Psychologists Press. 1991. P. 269–313.
- Borman W.C., Motowidlo S.J.* Expanding the Criterion Domain to Include Elements of Contextual Performance // *Schmitt N., Borman W.C. and Associates* (eds). *Personnel Selection in Organizations*. San Francisco: Jossey-Bass. 1993. P. 71–98.
- Egger M., Smith G.D., Schneider M., Minder C.* Bias in Meta-Analysis Detected by a Simple Graphical Test // *British Medical Journal*. 1997. Vol. 315. P. 629–634.
- Gaugler B.B., Rosenthal D.B., Thornton G.C., Bentson C.* Meta-Analysis of Assessment Center Validity // *Journal of Applied Psychology*. 1987. Vol. 72. P. 493–511.
- Hermelin E., Robertson I.T.* A Critique and Standardization of Meta-Analytic Validity Coefficients in Personnel Selection // *Journal of Occupational and Organizational Psychology*. 2001. Vol. 74. P. 253–277.
- Hunter J.E.* Test Validation for 12,000 Jobs: An application of job classification and validity generalization analysis to the general aptitude test battery // *USES Test Research Report*. 1983. No. 45, Division of Counseling and Test Development Employment and Training Administration, US Department of Labor, Washington, DC.
- Hunter J.E., Schmidt F.L.* *Methods of Meta-Analysis: Correcting error and bias in research findings*. Beverly Hills, CA: Sage. 1990.
- Hunter J.E., Schmidt F.L., Le H.* Implications of Direct and Indirect Range Restriction for Meta-Analysis Methods and Findings // *Journal of Applied Psychology*. 2006. Vol. 91. P. 594–612.
- McDaniel M.A., Morgeson F.P., Finnegan E.B., Campion M.A., Braverman E.P.* Use of Situational Judgment Tests to Predict Job Performance: A clarification of the literature // *Journal of Applied Psychology*. 2001. Vol. 86. P. 730–740.
- Roth P.L., Bobko P., Switzer F.S., Dean M.A.* Prior Selection Causes Biased Estimates of Standardized Ethnic Group Differences: Simulation and analysis // *Personnel Psychology*. 2001. Vol. 54. P. 591–617.
- Salgado J.F., Anderson N., Moscoso S., Bertua C., De Fruyt F., Rolland J.P.* A Meta-Analytic Study of General Mental Ability Validity for Different Occupations in the European Community // *Journal of Applied Psychology*. 2003. Vol. 88. P. 1068–1081.
- Schmidt F.L., Oh I., Le H.* Increasing the Accuracy of Corrections for Range Restriction: Implications for selection procedure validities and other research results // *Personnel Psychology*. 2006. Vol. 59. P. 281–305.
- Schmitt N., Gooding R.Z., Noe R.A., Kirsch M.* Meta-Analyses of Validity Studies Published Between 1964 and 1982 and the Investigation of Study Characteristics // *Personnel Psychology*. 1984. Vol. 37. P. 407–422.
- Thornton G.C., III.* *Assessment Centers and Human Resource Management*. Reading, MA: Addison-Wesley, 1992.
- Viswesvaran C., Ones D.S., Schmidt F.L.* Comparative Analysis of the Reliability of Job Performance Ratings // *Journal of Applied Psychology*. 1996. Vol. 81. P. 557–574.

### Список статей, послуживших основой для метаанализа

- Anderson L.R., Thaker J.* Self-Monitoring and Sex as Related to Assessment Center Ratings and Job Performance // *Basic and Applied Social Psychology*. 1985. Vol. 6. P. 345–361.
- Arthur W., Tubre T.* The Assessment Center Construct-Related Validity Paradox: An investigation of self monitoring as a misspecified construct, 2001. Unpublished Manuscript.
- Binning J.F., Adorno A.J., LeBreton J.M.* Intraorganizational Criterion-Based Moderators of



- Assessment Center Validity // Paper Presented at the Annual Conference of the Society for Industrial and Organizational Psychology. Atlanta, GA, 1999, April.
- Bobrow W., Leonards J.S.* Development and Validation of an Assessment Center During Organizational Change // *Journal of Social Behavior and Personality*. 1997. Vol. 12. P. 217–236.
- Burroughs W.A., White L.L.* Predicting Sales Performance // *Journal of Business and Psychology*. 1996. Vol. 11. P. 73–84.
- Chan D.* Criterion and Construct Validation of an Assessment Centre // *Journal of Occupational and Organizational Psychology*. 1996. Vol. 69. P. 167–181.
- Dayan K., Kasten R., Fox S.* Entry-Level Police Candidate Assessment Center: An efficient tool or a hammer to kill a fly? // *Personnel Psychology*. 2002. Vol. 55. P. 827–849.
- Dobson P., Williams A.* The Validation of the Selection of Male British Army Officers // *Journal of Occupational Psychology*. 1989. Vol. 62. P. 313–325.
- Feltham R.* Assessment Centre Decision Making: Judgmental vs. mechanical // *Journal of Occupational Psychology*. 1988. Vol. 61. P. 237–241.
- Fleenor J.W.* Constructs and Developmental Assessment Centers: Further troubling empirical findings // *Journal of Business and Psychology*. 1996. Vol. 10. P. 319–333.
- Fox S., Levonai-Hazak M., Hoffman M.* The Role of Biodata and Intelligence in the Predictive Validity of Assessment Centres // *International Journal of Selection and Assessment*. 1995. Vol. 3. P. 20–28.
- Goffin R.D., Rothstein M.G., Johnston N.G.* Personality Testing and the Assessment Center: Incremental validity for managerial selection // *Journal of Applied Psychology*. 1996. Vol. 81. P. 746–756.
- Goldstein H.W., Yusko K.P., Braverman E.P., Smith D.B., Chung B.* The Role of Cognitive Ability in the Subgroup Differences and Incremental Validity of Assessment Center Exercises // *Personnel Psychology*. 1998. Vol. 51. P. 357–374.
- Gomez J.J., Stephenson R.S.* Validity of an Assessment Center for the Selection of School-Level Administrators // *Educational Evaluation and Policy Analysis*. 1987. Vol. 9. P. 1–7.
- Higgs M.* The Value of Assessment Centres // *Selection and Development Review*. 1996. Vol. 12. P. 2–6.
- Hoffman C.C., Thornton G.C., III.* Examining Selection Utility Where Competing Predictors Differ in Adverse Impact // *Personnel Psychology*. 1997. Vol. 50. P. 455–470.
- Jones R.G., Whitmore M.D.* Evaluating Developmental Assessment Centers as Interventions // *Personnel Psychology*. 1995. Vol. 48. P. 377–388.
- McEvoy G.M., Beatty R.W.* Assessment Centres and Subordinate Appraisals of Managers: A seven year examination of predictive validity // *Personnel Psychology*. 1989. Vol. 42. P. 37–52.
- Moser K., Schuler H., Funke U.* The Moderating Effect of Raters' Opportunities to Observe Ratees' Job Performance on the Validity of an Assessment Centre // *International Journal of Selection and Assessment*. 1999. Vol. 7. P. 133–141.
- Nowack K.M.* Congruence Between Self–Other Ratings and Assessment Center Performance // *Journal of Social Behavior and Personality*. 1997. Vol. 12. P. 145–166.
- Pynes J., Bernardin H.J.* Entry-Level Police Selection: The assessment center is an alternative // *Journal of Criminal Justice*. 1992. Vol. 20. P. 41–55.
- Robertson I.* Predictive Validity of the General Fast Stream Selection Process. 1999. Unpublished Validity Report, School of Management, UMIST.
- Russell C.J., Domm D.R.* Two Field Tests of an Explanation of Assessment Centre Validity // *Journal of Occupational and Organizational Psychology*. 1995. Vol. 68. P. 25–47.
- Schmitt N., Schneider J.R., Cohen S.A.* Factors Affecting Validity of a Regionally Administered Assessment Center // *Personnel Psychology*. 1990. Vol. 43. P. 1–12.
- Thomas T., Sowinski D., Laganke J., Goudy K.* Is the Assessment Center Validity Paradox Illusory? // Paper Presented at the 20th Annual Conference of the Society for Industrial and

Organizational Psychology, Los Angeles, CA, 2005, April.

*Tziner A., Meir E.I., Dahan M., Birati A.* An Investigation of the Predictive Validity and Economic Utility of the Assessment Center for the High-Management Level // Canadian Journal of Behavioral Science. 1994. Vol. 26. P. 228–245.

## Приложение А

Таблица А1. Краткая сводка результатов исследований, включенных в метаанализ

Автор(ы)	Исходные общие коэффициенты валидности	Объем выборки (N)	Ограничение дисперсии критерия (U)	Коэффициенты валидности, скорректированные с учетом прямого ограничения дисперсии	Коэффициенты валидности, скорректированные с учетом прямого ограничения дисперсии и ненадежности критерия
Anderson, Thaker (1985)	.43	25	1	.43	.60
Artur, Tubre (2001)	.23	70	.94	.24	.34
Binning, Adorno, LeBreton (1999)	.15	1637	.8	.19	.26
Bobrow, Leonards (1997)	.23	71	1	.23	.32
Burroughs, White (1996)	.49	29	1	.49	.68
Chan (1996)	.06	46	1	.06	.08
Dayan, Kasten, Fox (2002)	.17	420	.94	.18	.25
Dobson, Williams (1989)	.14	450	.55	.25	.35
Feltham (1988)	.16	128	.52	.30	.41
Fleenor (1996)	.24	85	1	.24	.33
Fox, Levonai-Hazak, Hoffman (1995)	.33	91	1	.33	.46
Goffin, Rothstein, Johnson (1996)	.30	68	.88	.34	.47
Goldstein, Yusko, Braverman, Smith, Chung (1998)	.18	633	1	.18	.25
Gomez, Stephenson (1987)	.19	121	.94	.20	.28
Higgs (1996)	.32	123	.94	.34	.47
Hoffman, Thornton (1997)	.26	118	1	.26	.36
Jones, Whitmore (1995)	-.03	149	1	-.03	-.04
McEvoy, Beatty (1989)	.28	48	1	.28	.39
Moser, Schuler, Funke (1999)	.37	144	1	.37	.51
Nowack (1997)	.25	144	1	.25	.35
Pynes, Bernardin (1992)	.23	68	1	.23	.32
Robertson (1999)	.23	105	.94	.24	.34
Russell, Domm (1995)	.22	140	1	.22	.31
Russell, Domm (1995)	.23	172	1	.23	.32
Schmidt, Schneider, Cohen (1990)	.08	402	.95	.08	.12
Thomas, Sowinski, Laganke, Goudyn (2005, April)	.30	56	.94	.31	.43
Tziner, Meir, Dahan, Birati (1994)	.21	307	.94	.22	.31