

---

**ОРГАНИЗАЦИОННО-ПСИХОЛОГИЧЕСКАЯ ПРАКТИКА**

---

## **Дистанционная оценка управленческого потенциала: исследование критериальной и инкрементной валидности**

**ПОПОВ Алексей Юрьевич**

*Кандидат психологических наук, научный консультант «ЭКОПСИ Консалтинг», (Пермь)*

*E-mail: popov-research@mail.ru*

**ЛУРЬЕ Евгений Валентинович**

*Руководитель Центра разработки оценочных технологий «ЭКОПСИ Консалтинг», (Москва)*

*E-mail: lurie.eugene@gmail.com*

*В статье рассматриваются проблемы критериальной валидности дистанционных инструментов оценки, их отношение к диагностике и прогнозу управленческих компетенций и бизнес-результатам. В качестве кейса выступает масштабный проект, проведенный одним из крупнейших банков России совместно с компанией «ЭКОПСИ Консалтинг» в целях формирования кадрового резерва на должность руководителя начального звена. Дистанционная оценка выступала как один из важнейших этапов процедуры отбора и предшествовала проведению Центра оценки (ЦО). Четыре диагностических инструмента, использованных в дистанционной оценке, рассматриваются в статье как предикторы общего оценочного рейтинга в ЦО и далее – внешних показателей успешности на уровне отдельных работников и региональных подразделений банка. Результаты эмпирических исследований сопровождаются постановкой ряда проблем, связанных с необходимостью организации послепроектного сопровождения для такого рода оценочных проектов.*

**Ключевые слова:** дистанционная оценка; тестирование; инбаскет; прогноз результативности; валидность.

### **Проблема**

Оценка управленческих компетенций персонала заслуженно занимает лидирующее место среди практических задач организационной психологии. Для того чтобы организация функционировала и развивалась, необходимо, чтобы ключевые управленческие позиции были заняты профессионалами. Они должны уметь решать управленческие задачи, анализировать разноплановую информацию, быть готовыми к постоянному саморазвитию.

Вместе с тем результаты оценки этих качеств следует рассматривать не более чем как промежуточный этап работы. В конечном итоге, основная цель оценки управленческих компетенций персонала – принять такие кадровые решения, которые приведут к увеличению прибыли [Holling, 1998]. Выполнение этой задачи в силу ее сложности достигается в

оценочных проектах очень редко. Особенно острый дефицит наблюдается в работах российских специалистов, где эта задача только начинает ставиться на теоретическом уровне [Малыхин, 2004; Куприянов, 2011].

Если допустить, что управленческие компетенции оказывают влияние на успешность деятельности руководителя и уровень доходов организации, то инструменты оценки этих качеств должны быть связаны со следующими показателями:

- 1) успешность деятельности на индивидуальном уровне в настоящий момент;
- 2) успешность деятельности на индивидуальном уровне в будущем;
- 3) успешность деятельности на корпоративном уровне в настоящий момент;
- 4) успешность деятельности на корпоративном уровне в будущем.

Разумеется, возможны и возражения. Управленческие компетенции не обязательно должны быть прямо монетизируемыми. Если в настоящее время респондент не занимает управленческую должность, то успешность его деятельности в этой должности не обязательно будет коррелировать с управленческими компетенциями (просто потому, что он пока никем не управляет). Успешность деятельности на индивидуальном уровне не всегда просто измерить, а успешность деятельности на корпоративном уровне детерминируется множеством факторов, помимо управленческих компетенций руководителей.

При всех этих возражениях, однако, следует ожидать корреляций результатов выполнения инструментов оценки управленческих компетенций с указанными критериями на уровнях от слабого до умеренного [Хермелин и др., 2011]. Речь идет об обязательном и всестороннем исследовании критериальной валидности тех оценочных инструментов, которые используются для диагностики компетенций. Только в этом случае (при всей сложности и длительности подобных исследований) можно будет оценить, удалось ли организации с помощью разработанной оценочной технологии заработать деньги. А это, в конечном итоге, имеет для организационной психологии не только прикладной, но и фундаментальный смысл. Проблема заключается в том, что зачастую оценочные проекты ограничиваются лишь исследованием конструктивной валидности, а иногда и вовсе игнорируют подобные «затратные» работы.

Наиболее общепризнанным и валидным способом оценки управленческих компетенций являются Центры оценки (ЦО) [Lievens et al., 2009; Thornton, Rupp, 2006], однако их очевидным недостатком является цена. Это становится особенно актуальным, когда численность персонала, который необходимо направить на оценку, велика. Возникает вопрос: стоит ли тратить деньги на такое количество кандидатов, которые в конечном итоге не будут отобраны? Очевидно, напрашивается более оперативная, дешевая и вместе с тем валидная процедура, пригодная для первичного отбора кандидатов перед прохождением ЦО. Проблема заключается в том, чтобы доказать, что эта процедура, не являясь ЦО, тем не менее прогнозирует его возможные результаты с приемлемой степенью достоверности (достаточно, чтобы быть уверенным, что определенное число кандидатов можно отсеивать до прохождения ЦО, так как их оценка, скорее всего, не будет высокой).

В данной статье на примере конкретного проекта описывается опыт разработки процедуры комплексной дистанционной оценки, призванной служить «фильтром» для большого количества кандидатов перед прохождением ЦО. Особое внимание уделяется критериальной валидности дистанционной оценки по отношению к ЦО. Также описывается прогностический потенциал инструментов дистанционной оценки по отношению к внешним критериям успешности деятельности кандидатов.

Психометрический вызов, предъявляемый к такого рода процедуре, определяется тремя соображениями. Во-первых, она должна с достаточной степенью достоверности предсказывать уровень развития компетенций у кандидата, не измеряя их напрямую (т.е. не оценивая непосредственно наблюдаемое поведение). Во-вторых, результаты дистанцион-

ной оценки должны быть связанными с внешними объективными показателями успешности деятельности кандидата. В-третьих, для обеспечения валидности инструменты оценки должны быть достаточно разнообразными (по форме и по содержанию), но вместе с тем не должны быть избыточными.

### **Критериальная и конструктивная валидность оценочных инструментов**

Исследование прогностической способности результатов оценки персонала по отношению к успешности деятельности на языке психометрики называется исследованием критериальной валидности. В данном случае результаты оценки используются как «предикторы», а показатели успешности деятельности (например, суммарная прибыль компании) – как та зависимая переменная, значение которой пытаются спрогнозировать [Фэр, Бакаррак, 2010].

Однако критериальная валидность является не единственным показателем качества оценочных инструментов. Помимо нее, необходимо исследовать и конструктивную валидность: измеряют ли инструменты именно то, что они призваны измерять, либо же нечто другое? [Фэр, Бакаррак, 2010].

Здесь следует отметить, что отношения между конструктивной и критериальной валидностью неоднозначны, и каждый исследователь вынужден формировать собственную позицию относительно того, какой из видов валидности будет считаться приоритетным [Christian et al., 2010]. На примере задачи формирования управленческого кадрового резерва представим наиболее типичные результаты исследований:

1. Конструктивная валидность инструментов оценки оказалась высокой: доказано, что инструменты измеряют именно управленческие компетенции (либо же их предпосылки). Однако критериальная валидность инструментов низка: баллы, заработанные кандидатами по результатам оценки, плохо прогнозируют их успешность в профессиональной деятельности.
2. Низкая конструктивная валидность инструментов оценки на фоне высокой критериальной валидности. Другими словами, мы не понимаем, что именно мы измерили, но это «что-то» хорошо прогнозирует успешность руководителей [Christian et al., 2010]. С одной стороны, это позитивный знак, а с другой – непонимание конструкта чревато множеством артефактов и неправильных толкований результатов исследования.

С этой дилеммой сталкиваются все специалисты по оценке. Текущий уровень развития науки позволяет без особых усилий получить воспроизводимые инструменты, хорошо работающие на критериальную валидность. В первую очередь это обычные тесты способностей. Однако конструктивная валидность этих инструментов применительно к оценке управленческих компетенций вызывает очевидные сомнения.

Борьба за конструктивную валидность часто оказывается благородным, но бесполезным для практики порывом. Инструменты приобретают внутреннюю «стройность» и чистоту, но становятся подвержены множеству дополнительных помех в ситуациях экспертизы. В результате связи с внешними критериями ослабевают, и это приводит к необоснованности проведения отбора на основании проведенной оценки. В связи с этим нашей задачей в описываемом проекте являлось обеспечение достаточного уровня конструктивной валидности с последующим исследованием и повышением критериальной валидности.

### **Проект дистанционной оценки: описание и инструменты**

Оценочный проект, описываемый в данной статье, преследовал задачу формирования кадрового резерва на должность руководителя дополнительного офиса (ДО) в банке. Перед

организацией встала задача разработать систему отбора и подготовки рядовых сотрудников к занятию этой ключевой начальной управленческой позиции (всего в банке несколько тысяч ДО). В круг возможных кандидатов были включены как рядовые сотрудники ДО (операционисты, кассиры, кредитные специалисты), так и административные работники (заместители руководителей ДО, руководители секторов ДО).

Для целевой позиции была разработана модель компетенций, которая включала следующие составляющие: управление командой и операционной деятельностью, работа с клиентами, аналитика, коммуникация и влияние, управленческая ответственность.

Отбор осуществлялся в четыре этапа по логике «воронки»:

- отбор по формальным показателям – образование, опыт и результаты работы;
- отбор по мотивации – готовность сотрудника развиваться и желание выполнять обязанности руководителя ДО;
- дистанционная оценка управленческого потенциала (см. ниже);
- Центр оценки – оценка управленческих компетенций.

Первый этап был проведен сотрудниками банка с помощью анализа кадровой базы данных. В пул кандидатов были включены сотрудники, рекомендованные руководителями в резерв, а также сотрудники-самовыдвиженцы (всего более 21 000 человек). Второй этап проводился с помощью анкетирования отобранных сотрудников. В анкете содержались прямые вопросы относительно готовности сотрудника прикладывать дополнительные усилия для попадания в кадровый резерв и желания работать на должности руководителя ДО.

Ключевым этапом проекта стала дистанционная оценка, и именно анализу третьего этапа посвящена данная статья. На этой стадии отбора приняли участие 5257 сотрудников банка. Средний возраст участников – 33.6 года, стандартное отклонение – 6.6 года. Средний опыт работы в финансовой сфере – 7 лет 11 месяцев (стандартное отклонение – 6 лет). Средний опыт работы в данном банке – 7 лет 6 месяцев (стандартное отклонение – 5 лет 9 месяцев). Дистанционная оценка предполагала трехчасовое тестирование в онлайн-режиме и работу с четырьмя инструментами, три из которых были разработаны специально для данной оценки.

1. *Тест знаний в области банковского бизнеса («Профтест»)* – измеряет общие знания в сфере банковской деятельности, а также знания о работе ДО банка. Как и любой тест знаний, этот инструмент имеет четко выраженные правильные и неправильные ответы.
2. *Управленческий «Кейс-тест»* – измеряет неявные знания и установки в области управления бизнес-задачами и организации работы с клиентами. По форме представляет собой набор ситуаций и четыре варианта ответа к каждой, описанных в виде текста.
3. *Моделирующее упражнение «Электронный инбаскет»* – оценивает умение принимать решения на основе разрозненной информации и планировать реализацию управленческих задач. Упражнение имитирует работу с электронной почтой и календарем. Помимо компьютерного формата проведения, от традиционных инбаскетов отличается частичной предзаданностью вариантов действий, необходимой для автоматизированной обработки результатов.
4. *Краткий тест отбора КТО-2* (разработчик – лаборатория «Гуманитарные технологии»)¹ – предназначен для измерения общих интеллектуальных способностей, включает лексические и вычислительные задания, задания на эрудицию и внимание. Основными преимуществами этого теста перед другими методиками

1 Описание теста см. по ссылке <http://www.ht.ru/cms/component/content/article/101078> (дата обращения: 10.03.2012).

диагностики интеллектуальных способностей являются быстрота и легкость проведения. Наличие расширенного банка заданий и их рандомизация в ходе тестирования надежно защищают инструмент от фальсификации, что критически важно в ситуации отбора.

По замыслу данная батарея методик (вместе с предварительной мотивационной анкетой) покрывала широкий спектр значимых деловых качеств, выявленных в результате анализа работ на первом этапе разработки. Явные и неявные знания о работе в банке измерялись в «Профтесте» и «Кейс-тесте» соответственно, общие способности – в КТО-2, мотивационные компоненты оценивались в анкете. «Электронный инбаскет» был призван измерять как знания и способности, так и отдельные поведенческие характеристики, приближающиеся по существу к компетенциям [Lievens et al., 2010].

Очевидно, что результаты дистанционной оценки не измеряют напрямую сами управленческие компетенции. Причина в том, что в ходе онлайн-тестирования не оценивается непосредственно наблюдаемое поведение (максимум того, что оценивается, – это продукты этого поведения). Однако набор критериев, измеряемых внутри батареи (см. выше), был сконструирован таким образом, чтобы отразить существенные компоненты компетенций руководителя ДО (интеллектуальные, мотивационные, коммуникационные и знаниевые). Таким образом, на уровне критериев можно говорить о комплексной оценке «потенциала развития компетенций» [Bartram, 2005].

Для оценки критериев, выявленных в ходе анализа должности, были созданы специальные методики. Две из них – «Кейс-тест» и «Электронный инбаскет» – были разработаны на основе принципов моделирования управленческих ситуаций, используемых также в Центрах оценки [Guidelines..., 2009]<sup>2</sup>. Такой подход с применением моделирующих упражнений давал дополнительные основания полагать, что указанные инструменты могли быть использованы для достоверного прогноза уровня развития управленческих компетенций кандидатов. Это и определило логику построения проекта по типу «воронки» и задачу дистанционной оценки в ней. Задача состояла в существенном сокращении числа оцениваемых кандидатов для пропуска на очные ЦО при обеспечении высокого качества отбора.

### **Критерии успешности**

В описываемом исследовании результаты выполнения четырех методик дистанционной оценки выступали как предикторы (независимые переменные, на основе значений которых делается попытка предсказать значения объективных показателей результативности). В качестве внешних критериев выступали следующие показатели:

- фактический коэффициент квартального премирования;
- оценка работника руководителем по специальному универсальному набору критериев оценки;
- рейтинг региональных подразделений банка по бизнес-показателям (далее по тексту – рейтинг подразделений);
- общий оценочный рейтинг (ООР) в Центрах оценки (ЦО).

Фактический коэффициент квартального премирования – это показатель, характеризующий степень выполнения сотрудником индивидуального плана за квартал. Оценка руководителем – это пятиуровневая оценка, проводимая раз в квартал и содержащая пять универсальных критериев, связанных с ключевыми ценностями банка (ориентация на клиента,

<sup>2</sup> Перевод данной статьи опубликован в журнале «Организационная психология»: Нормативы и этические принципы создания и проведения Центра оценки [Электронный ресурс] // Организационная психология. 2011. Т. 1. №2. С. 26–44. URL: www.orgpsyjournal.hse.ru – Прим. ред.

командная работа и т.п.). Важно, что в процедуре объединены экспертная оценка, выставляемая руководителями подчиненным, и самооценка.

Для расчета бизнес-показателей были использованы данные по 17 региональным подразделениям. По каждому из них имелись данные о финансовых показателях за три квартала. Данные были сведены в единый рейтинг подразделений.

ЦО проводились для тех сотрудников, которые успешно прошли все предыдущие этапы отбора. Участники оценивались по 5 управленческим компетенциям в ходе однодневного мероприятия. В процедуре использовались как групповые, так и индивидуальные упражнения. Оценку проводили подготовленные наблюдатели-эксперты. После завершения каждой оценочной сессии группа наблюдателей проводила согласование оценок по каждому критерию.

### **Обсуждение критериев успешности**

Поиск оптимальных критериев является одной из самых острых проблем психологии. Как видно, ни один из приведенных критериев не является идеальным. Например, результаты ЦО не могут выступать объективным показателем результативности. Это «всего лишь» экспертная оценка управленческих компетенций независимыми наблюдателями. Вместе с тем ЦО конструируются по принципу «множественности» (критериев, наблюдателей, упражнений) [Guidelines..., 2009]. При соблюдении необходимых стандартов разработки и проведения с их помощью можно получать информативные и качественные данные, позволяющие прогнозировать успешность [Хемерлин и др., 2011; Сергиенко, 2000]. Наиболее существенным является то, что ЦО позволяют оценивать поведение, т.е. собственно компетенции в их непосредственном выражении [Lievens et al., 2009]. На сегодняшний день существует достаточно оснований предполагать, что общий оценочный рейтинг по результатам ЦО является одним из наиболее точных предикторов управленческого потенциала [Thornton, Rupp, 2006].

Особый интерес ЦО для исследования состоит в том, что его результаты могут выступать как показатель и конструктивной, и критериальной валидности дистанционной оценки. Анализировать конструктивную валидность возможно, так как в ЦО измеряется несколько переменных (управленческих компетенций). Взаимосвязь баллов дистанционной оценки и результатов по различным компетенциям позволяет исследовать конструктивное «наполнение» разработанных инструментов. Оценка критериальной валидности возможна, поскольку предполагается, что уровень развития управленческих компетенций прогнозирует успешность руководителя.

Регулярная оценка работника руководителем выступает важным показателем результативности. Она используется во всем банке и является достаточно универсальным мерилем, позволяющим сравнивать людей из разных регионов и подразделений. Однако, во-первых, эта оценка отражает текущие результаты сотрудников, а чаще всего оценку проходили рядовые сотрудники. Нас же больше всего интересовали управленческие критерии. Во-вторых, с психометрической точки зрения этот вид оценки имеет определенные ограничения. Причина в том, что данная оценка является экспертной. Источником оценок здесь выступает непосредственный руководитель работника, следовательно, объективность контролировать непросто (особенно в масштабах крупной организации). Немаловажным является и то, что стандартизация процедуры оценки проходит этап становления: в различных подразделениях критерии и процедура оценки могут не быть полностью идентичными.

На первый взгляд объективным показателем кажется фактический коэффициент квартального премирования сотрудника. Однако анализ данных показал, что в исследуемой выборке преобладает значение 1 (сто процентное выполнение плана). Это означает, что фактический коэффициент квартального премирования обладает низкой дискриминатив-

ностью. Несмотря на указанные недостатки, использование этих критериев в комплексе, вероятно, даст более объективную картину.

Отдельного внимания заслуживает рейтинг подразделений банка по финансовым показателям, который, по нашему мнению, соответствует высоким стандартам объективного критерия. В нем усредняются отдельные случайные погрешности, связанные с ошибкой измерения. Сопоставление рейтингов подразделений по дистанционной оценке с финансовым рейтингом этих регионов может дать надежные оценки критериальной валидности оценочных инструментов. Недостатком этого критерия является его «неперсонализированность» – мы располагаем лишь общим результатом по региону и знаем его относительное место в списке регионов. Финансовых результатов по отдельным участникам нет (что понятно). В связи с этим для расчетов использовался усредненный результат всех участников из одного региона.

### **Результаты исследования критериальной валидности**

Взаимосвязь результатов выполнения четырех инструментов дистанционной оценки с описываемыми критериями приведены в табл. 1. Пояснения требует тот факт, что при исследовании критериальной валидности инструментов дистанционной оценки в одних случаях используются коэффициенты корреляции суммарного балла и критерия («Профтест» и КТО-2), в других – коэффициенты детерминации в уравнениях множественной регрессии, где в качестве независимых переменных выступают отдельные пункты или части диагностической процедуры («Кейс-тест» и «Электронный инбаскет»). Аргументами в пользу применения такого подхода были следующие:

- 1) применение уравнений множественной регрессии, где в качестве независимых переменных выступали бы отдельные пункты «Профтеста» или КТО-2, невозможно, поскольку пункты данных тестов предъявлялись респондентам случайным образом из банка пунктов, следствием чего явилось большое количество пропущенных данных в матрице сырых баллов;
- 2) инструменты типа «Кейс-теста» и «Электронного инбаскета», в отличие от тестов знаний и способностей, известны своей «мультиконструктностью» [Lievens et al., 2008; Lievens et al., 2010]. Наши исследования в рамках данного проекта также подтверждают это положение. В частности, усредненная межпунктовая корреляция между 13 ситуациями, входящими в «Кейс-тест», составляет 0.16, что практически свидетельствует о слабой зависимости измерений друг от друга и, таким образом, соответствует требованиям множественной регрессии;
- 3) количество отдельных «пунктов» той или иной измерительной процедуры, входящих в состав уравнения регрессии в качестве независимых переменных, не было слишком велико. Для «Кейс-теста» это были 13 ситуаций, для «Электронного инбаскета» – только те первичные действия пользователя при работе с упражнением, которые являются обязательными при любом варианте развертывания сценария;
- 4) для того чтобы избежать исследовательских артефактов, при расчете коэффициентов детерминации в регрессионных уравнениях по возможности применялась логика кросс-валидации. Для этого выборка случайным образом делилась на две части – тренировочную и тестовую, на тренировочной подвыборке рассчитывались регрессионное уравнение и все бета-коэффициенты, которые затем без модификаций перепроверялись на тестовой подвыборке. За итоговый результат принимался коэффициент детерминации, полученный именно на тестовой подвыборке [Picard, Cook, 1984].

Вместе с тем следует осознавать и ограничение в применении такого подхода. Коэффициенты корреляции между критерием и суммарным баллом по тесту (с одной стороны) и между критерием и совокупностью первичных параметров/ситуаций (с другой стороны) очевидно различны. В первом случае результат по тесту высчитывается на основе предположения о равенстве весов отдельных показателей в составе суммарного балла, во втором случае это взвешенная сумма. При этом, хотя взвешенная сумма и может быть ближе к реальности, размер весов определялся на основе эмпирических критериев, не получивших теоретического обоснования. Как следствие, корреляция между критерием и суммарным баллом может использоваться в целях индивидуальной интерпретации и обратной связи, в то время как результаты множественной регрессии – нет. Исключение составляет регрессия «Электронного инбаскета» на результаты ЦО. Причина в том, что на основе этого регрессионного уравнения вычисляется итоговый балл самого инструмента, который предоставляется клиенту. Во всех остальных случаях коэффициенты детерминации отдельных пунктов «Кейс-теста» и «Электронного инбаскета» с внешними критериями выражают скорее потенциал этих инструментов в предсказании внешних критериев, но не реально действующую возможность.

**Таблица 1.** Взаимосвязь результатов дистанционной оценки с внешними критериями

Инструмент дистанционной оценки	Оценка руководителем	Фактический коэффициент квартального премирования	Рейтинг подразделений	ЦО (общий оценочный рейтинг)
«Электронный инбаскет»	$(R^2)^1 = 0.22$ $F(87.2192) = 1.18,$ $p = 0.13$ ( $N = 1105$ )	$(R^2)^1 = 0.25$ $F(87.2192) = 1.61,$ $p = 0.000$ ( $N = 1882$ )	$r^2 = 0.19$	$r^3 = 0.61^{***}$ ( $N = 178$ )
«Профтест»	$r^3 = -0.007$ ( $N = 2662$ )	$r^3 = -0.004$ ( $N = 1985$ )	$r^2 = 0.31$	$r^3 = -0.01$ ( $N = 166$ )
«Кейс-тест»	$(R^2)^1 = 0.17$ $F(13.566) = 1.3,$ $p = 0.206$ ( $N = 567$ )	$(R^2)^1 = 0.24$ $F(13.276) = 1.3,$ $p = 0.204$ ( $N = 277$ )	$r^2 = 0.31$	$r^3 = 0.12$ ( $N = 158$ )
КТО-2	$r^3 = -0.015$ ( $N = 2573$ )	$r^3 = 0.074^{***}$ ( $N = 1973$ )	$r^2 = 0.57$	$r^3 = 0.18^*$ ( $N = 160$ )

<sup>1</sup> Коэффициент детерминации в уравнении множественной регрессии. Приведены результаты исследования статистической достоверности соответствующего коэффициента детерминации в однофакторном дисперсионном анализе ANOVA. Показатель  $F$  выражает величину эффекта,  $p$  – уровень статистической достоверности. В скобках приведены количество степеней свободы, связанное с регрессией, и общее количество степеней свободы.  $N$  – количество респондентов, принимавших участие в исследовании.

<sup>2</sup> Коэффициенты корреляции Спирмена между рейтингом территориальных подразделений банка по финансовым показателям и усредненным баллом, набранным сотрудниками по соответствующему инструменту оценки.

<sup>3</sup> Коэффициент корреляции Пирсона. В скобках указан размер выборки.

\*  $p < 0.05$ , \*\*\*  $p < 0.001$ .

### Обсуждение результатов исследования критериальной валидности

1. В целом результаты дистанционной оценки хорошо прогнозируют общий оценочный рейтинг в ЦО. При этом в данном прогнозе наибольшее значение имеют показатели «Электронного инбаскета». Это неудивительно, поскольку сама его разработка была «заточена» на то, чтобы максимально приблизиться к наиболее вероятным оценкам в ЦО. И вместе с тем сама процедура упражнения «Электронного инбаскета» предполагает решение ряда управленческих задач в моделируемой реальности. Таким образом, несмотря на частичную заданность вариантов действий, «Электронный инбаскет» по смыслу ближе



к ЦО, чем остальные инструменты дистанционной оценки [Schippmann et al., 1990; Lievens, Anseel, 2007; Lievens et al., 2010]. Умеренные значения в предсказании общего оценочного рейтинга имеют также «Кейс-тест» и КТО-2. «Кейс-тест» диагностирует неявные знания респондента в области того, какие управленческие решения являются более эффективными [Wagner, Sternberg, 1991]. Вероятно, такие знания являются предпосылкой эффективного поведения руководителя – необходимой, но не достаточной. То же самое можно сказать и об общих интеллектуальных способностях, диагностируемых в КТО-2. Таким образом, максимальный эффект предполагается от объединения этих инструментов.

2. Критериальную валидность разработанного «Кейс-теста» можно сравнить с аналогичными проектами, обобщенными в современных метаанализах [McDaniel et al., 2007]. Кейс-тесты обычно демонстрируют связь с внешними объективными критериями на уровне 0.20 с 80%-ным интервалом доверия от 0.13 до 0.38. Это справедливо для кейсовых инструментов со «знаниевой» инструкцией. Ф. Лайвенс дает среднюю оценку 0.34 [Lievens et al., 2008]. В целом результаты исследования (хотя нами и использовалось регрессионное уравнение с применением отдельных пунктов, а не суммарный балл по тесту) дают понять, что управленческий «Кейс-тест» в его сегодняшнем виде характеризуется достаточной критериальной валидностью по отношению к оценке результативности работника в его текущей должности.

3. Результаты выполнения теста профессиональных знаний в банковской сфере («Профтест») не предсказывают общий оценочный рейтинг в ЦО, но взаимосвязаны (на уровне усредненных данных) с рейтингом подразделений. Одно из возможных объяснений этого факта заключается в том, что эффективность деятельности сотрудников компании зависит не только от управленческих компетенций (при всей их важности), но и от простых знаний и эрудиции в своей профессиональной области (подкрепляется это рассуждение и высоким коэффициентом корреляции между рейтингом подразделений и результатами выполнения интеллектуального теста). Второе возможное объяснение заключается в «переворачивании» направления корреляционной связи. Вероятно, в подразделениях с высоким рейтингом (предположительно имеющих в среднем более компетентных сотрудников) кадровая политика осуществляется таким образом, что при отборе «новичков» уделяется больше внимания интеллектуальным способностям и общим познаниям в сфере банковского бизнеса.

4. Такие показатели индивидуальной результативности, как оценка руководителем и фактический коэффициент квартального премирования, вероятно, не прогнозируются результатами дистанционной оценки напрямую. Вместе с тем потенциал для такого прогноза в них уже имеется. С одной стороны, по логике построения данного оценочного проекта дистанционная оценка представляет собой промежуточный «фильтр» перед оценкой кандидатов в ЦО. То есть ее основная цель – не спрогнозировать результативность индивидуальной профессиональной деятельности, а отсеять тех кандидатов, которые (скорее всего) не получат высокого оценочного рейтинга в ЦО. С другой стороны, при использовании регрессионных уравнений обнаруживается прогностический потенциал инструментов дистанционной оценки: уже эти результаты могут объяснять до 6% разброса данных в фактическом коэффициенте квартального премирования и в оценке руководителей, что в социальных науках считается умеренным показателем [Cohen, 1988]. К сожалению, для того, чтобы оценить взаимосвязь общего оценочного рейтинга (ЦО) и внешних показателей результативности, в проекте еще не накоплено достаточно данных.

5. Рейтинг подразделений по бизнес-показателям наилучшим образом предсказывается общими интеллектуальными способностями (тест КТО-2). Это согласуется с данными многочисленных научных исследований, в которых показано, что общие интеллектуальные способности демонстрируют высокие показатели критериальной валидности [Schmidt, Hunter, 1998]. Полученный результат относится к описанной выше дилемме – высокая крите-

риальная валидность не всегда означает высокую конструктивную валидность инструмента. Управленческий потенциал, безусловно, требует определенного уровня развития общих интеллектуальных способностей. Однако валидность заданий такого теста по содержанию не в полной мере отвечает содержанию управленческой деятельности. Невысокие показатели связи КТО-2 и ЦО подтверждают тезис о том, что конструктивная валидность такого рода инструментов по отношению к оценке управленцев существенно ниже критериальной.

### Инкрементная валидность инструментов дистанционной оценки

Особую важность в такого рода проектах имеет вопрос о том, являются ли использованные инструменты избыточными [Lievens et al., 2008]. Этот вопрос имеет прямое отношение к максимальной результативности оценочного проекта (минимум стоимости, максимум эффекта). Для эмпирического ответа на него нами исследовались показатели инкрементной валидности инструментов, включенных в дистанционную оценку. Прежде всего, важно было исследовать инкрементную валидность «Электронного инбаскета» (способность к предсказанию профессиональной успешности помимо той доли информации, которая объясняется другими инструментами, проводившимися одновременно с «Электронным инбаскетом»), поскольку введение в общий оценочный «пакет» такой длительной и громоздкой процедуры требует не только теоретического (экспертного), но и эмпирического обоснования. Соответствующие показатели приведены в табл. 2.

**Таблица 2.** Инкрементная валидность «Электронного инбаскета»

По отношению к:	Коэффициент детерминации R <sup>2*</sup>	
	«Профтест», «Кейс-тест», КТО (первый шаг иерархической регрессии)	«Электронный инбаскет» (второй шаг иерархической регрессии)
Суммарная оценка наблюдателями в ЦО	.20	.63
Оценка руководителя	.03	.22
Фактический коэффициент квартального премирования	.08	.26

\* Использовалась иерархическая множественная регрессия. На первом шаге в модель вводились показатели «Профтеста», «Кейс-теста» и КТО-2, на втором шаге – пункты упражнения «Электронный инбаскет». Рост значений в коэффициенте детерминации от первого шага ко второму выражает степень «прироста» в доле дисперсии, которую «Электронный инбаскет» объясняет в том или ином критерии, помимо того, что уже объясняется тремя другими инструментами.

Основной вывод, который можно сделать по результатам данных табл. 2, таков: «Электронный инбаскет» действительно привносит важную долю информации в предсказание успешности деятельности кандидата, следовательно, является важной частью общей процедуры оценки. Умение принимать решения на основе разрозненной информации и планировать реализацию управленческих задач оказывается неотъемлемой частью комплексной диагностики управленческого потенциала. В наибольшей степени это очевидно по отношению к суммарной оценке наблюдателями в ЦО (прирост в коэффициенте детерминации от 0.2 до 0.63), однако заметно выражено и для других показателей.

Особый интерес представляла также инкрементная валидность «Кейс-теста» по отношению к КТО-2. В исследованиях, в том числе в метааналитических, показано, что кейсовые инструменты обладают некоторой инкрементной валидностью по отношению как к тестам интеллекта, так и к личностным опросникам. В частности, инкрементная валидность по отношению к тестам интеллекта для кейсовых методик составляет от 0.03 до 0.05

[McDaniel et al., 2007]. Это небольшая величина, однако исследователи признают ее важной и интерпретируют как свидетельство того, что кейсовые методики предоставляют важную информацию *помимо* общего интеллектуального уровня кандидата.

Итак, в качестве предикторов использовались балл по тесту КТО-2 (вводился в иерархическое регрессионное уравнение первым шагом) и 13 пунктов-ситуаций «Кейс-теста» (вводились в регрессионное уравнение вторым шагом). Результат: коэффициент детерминации в предсказании критерия для КТО-2 составляет 0.072, коэффициент детерминации при введении в иерархическое уравнение на втором шаге пунктов-ситуаций управленческого «Кейс-теста» составляет 0.249. Таким образом, по отношению к данному критерию инкрементная валидность теста составляет 0.177 и превышает данные, приведенные в научной литературе.

Впрочем, при использовании в регрессионном уравнении не 13 отдельных ситуаций, а суммарного балла по «Кейс-тесту» показатели инкрементной валидности снижаются. Коэффициент детерминации в предсказании критерия для КТО-2 в этом случае составляет 0.053, коэффициент детерминации при введении в иерархическое уравнение на втором шаге суммарного балла по «Кейс-тесту» равен 0.054.

Как можно проинтерпретировать такой результат? С одной стороны, инкрементная валидность суммарного балла по «Кейс-тесту» по отношению к общим интеллектуальным способностям (КТО-2) низка – она составляет 0.01. С другой стороны, соответствующие показатели в западных исследованиях также невысоки (0.03–0.05), однако интерпретируются как важные и значимые [McDaniel et al., 2007]. На текущем этапе исследования может быть предложена следующая трактовка результатов: неявные знания в области управления бизнес-задачами и организации работы с клиентами, измеряемые «Кейс-тестом», связаны с успешностью деятельности не напрямую, а через опосредующее звено – управленческие компетенции. Другими словами, чтобы влиять на успешность решения профессиональных задач, неявное должно, вероятно, стать проявленным поведением. С этой точки зрения использование «Кейс-теста» в качестве промежуточного этапа перед проведением ЦО в нашем проекте представляется правильным.

Впрочем, как уже было сказано, инкрементная валидность «Кейс-теста» возрастает при использовании не суммарного балла, а 13 отдельных ситуаций, входящих в его состав. Мультиконструктность, присущая кейс-тестам [Lievens et al., 2008], позволяет предполагать, что этот факт может являться не просто статистическим феноменом, но и иметь осмысленное объяснение с точки зрения теории. Различные стороны и аспекты неявных знаний в сфере управления могут вносить разный вклад в успешность деятельности. Однако нахождение этого объяснения – дело дальнейших самостоятельных исследований. В любом случае на настоящий момент можно утверждать, что результаты исследования инкрементной валидности «Кейс-теста» по отношению к внешним критериям успешности деятельности (доступным в данном проекте) могут использоваться лишь в сугубо исследовательских целях для изучения *потенциала* «Кейс-теста» в прогнозе такого рода критериев.

## Выводы

1. Задачи дистанционной оценки были достигнуты. По результатам психометрического исследования есть основания утверждать, что данные разработанной комплексной батареи методик (особенно «Электронный инбаскет», КТО-2 и «Кейс-тест») достоверно прогнозируют уровень развития управленческих компетенций, оцениваемых в Центрах оценки. Сама дистанционная оценка в предложенной компоновке инструментов может использоваться как способ отсева кандидатов перед более дорогостоящими этапами оценки.

2. Результаты дистанционной оценки обладают прогностическим потенциалом по отношению к таким показателям результативности отдельно взятых сотрудников, как оценка руководителя и фактический коэффициент квартального премирования. Вместе с тем есть основания полагать, что этот потенциал может быть выше, если усовершенствовать статистические характеристики показателей результативности, а также использовать *будущую* результативность вместо текущей.
3. На уровне усредненных данных (рейтинг подразделений) прогностический потенциал инструментов дистанционной оценки может быть назван высоким. На общем фоне особенно выделяются результаты классического теста способностей (КТО-2), корреляция которых с финансовыми результатами региональных подразделений является еще одним доказательством значимости подобных инструментов в дистанционной оценке персонала.
4. Один из ключевых параметров – корреляция общего оценочного рейтинга (ЦО) с показателями результативности – в данном исследовании не оценивался по причине недостаточности данных.

Кроме того, результаты проекта приводят к нескольким важным соображениям, относящимся ко всем оценочным проектам, преследующим задачу формирования кадрового резерва.

1. Важность послепроектного сопровождения в такого рода исследованиях трудно переоценить. Под послепроектным сопровождением понимается: а) отслеживание профессиональной результативности и карьерного развития сотрудников, прошедших оценку; б) периодический пересмотр эмпирических ключей и обновление норм; в) накопление эмпирических данных, которые бы позволили более комплексно оценить место управленческого потенциала в прогнозе успешности деятельности предприятия. Ни одно измерение не является идеальным, но в известной мере присущие измерению погрешности можно «победить», реализуя комплексную исследовательскую программу.
2. Не менее важны поиск и разработка надежных и достоверных показателей результативности деятельности сотрудников. С сожалением приходится констатировать, что практика сбора подобных критериальных мер развита в России слабо (как в среде специалистов-консультантов, так и в HR-сообществе). Полагаем, что для всесторонних исследований и для мониторинга качества использования результатов оценки персонала необходимо применять показатели результативности на уровне как отдельных работников, так и структурных подразделений. Причем в обоих случаях необходима комплексная оценка. На уровне отдельных работников речь может идти об объединении экспертных оценок (например, оценка руководителями) и объективных показателей результативности. Однако и одно, и другое теряет свой психометрический и исследовательский смысл, если является недискриминативным. Следовательно, психологам необходимы регулярные статистические исследования самих показателей результативности предприятия. На уровне структурных подразделений речь может идти об общей прибыли за определенный промежуток времени.

### Список литературы

Куприянов Е.А. Стоит ли игра свеч: Валидность Центров оценки [Электронный ресурс] // Организационная психология. 2011. Т. 1. №1. С. 50–58. URL: <http://orgpsyjournal.hse.ru> (дата обращения: 01.02.2012).

- Малыхин П.А.* Ассесмент-Центр – дорогое удовольствие? // Персонал-Микс. 2004. №7. С. 45–48.
- Сергиенко С.К.* Групповые технологии решения организационно-психологических задач: Автореф. дисс. ... докт. психол. наук. М.: ГУУ, 2000.
- Фер М., Бакарак В.* Психометрика: введение / Пер. с англ. под ред. Н.А. Батурина, Е.В. Эйдмана. Челябинск: Издательский центр ЮУрГУ, 2010.
- Хермелин И., Лайвенс Ф., Робертсон А.* Прогностическая валидность Центра оценки: мета-анализ / пер. с англ. Е. Лурье, В. Тихомирова [Электронный ресурс] // Организационная психология. 2011. Т. 1. №1. С. 39–49. URL: <http://orgpsyjournal.hse.ru> (дата обращения: 01.02.2012).
- Bartram D.* The Great Eight Competencies: A criterion-centric approach to validation // *Journal of Applied Psychology*. 2005. Vol. 90. Is. 6. P. 1185–1203.
- Bray G.W., Grant D.L.* The Assessment Center in the Measurement of Potential for Business Management // *Psychological Monographs: General and Applied*. 1966. Vol. 80. Is. 17. P. 1–27.
- Boudreau J.W., Ramstad P.M.* Strategic I/O Psychology and the Role of Utility Analysis Models [Электронный ресурс] // CAHRS Working Paper Series. Paper 57. 2002. URL: <http://digitalcommons.ilr.cornell.edu/cahrswp/57> (дата обращения: 19.02.2012).
- Christian M.S., Edwards, B.D., Bradley J.C.* Situational Judgment Tests: Constructs Assessed and a Meta-analysis of their Criterion-related Validities // *Personnel Psychology*. 2010. Vol. 63. Is. 1. P. 83–117.
- Cohen J.* Statistical Power Analysis for the Behavioral Sciences (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 1988.
- Guidelines and Ethical Considerations for Assessment Center Operations // *International Journal of Selection and Assessment*. 2009. Vol. 17. Is. 3. P. 243–253.
- Holling H.* Utility Analysis of Personnel Selection: An Overview and Empirical Study Based on Objective Performance Measures // *Methods of Psychological Research Online*. 1998. Vol. 3. Is. 1. P. 5–24.
- Lievens F., Anseel F.* Creating alternate in-basket forms through cloning: Some preliminary results // *International Journal of Selection and Assessment*. 2007. Vol. 15. Is. 4. P. 405–411.
- Lievens F., Peeters H., Schollaert E.* Situational Judgment Tests: A Review of Recent Research // *Personnel Review*. 2008. Vol. 37. Is. 4. P. 426–441.
- Lievens F., Tett R.P., Schleicher D.J.* Assessment Centers at the Crossroads: Toward a Reconceptualization of Assessment Center Exercises // J.J. Martocchio, H. Liao (eds.). *Research in Personnel and Human Resources Management*. 2009. Bingley: JAI Press. P. 99–152.
- Lievens F., Van Keer E., Volckaert E.* Gathering Behavioral Samples Through a Computerized and Standardized Assessment Center Exercise: Yes, It Is Possible // *Journal of Personnel Psychology*. 2010. Vol. 9. Is. 2. P. 94–98.
- McDaniel M.A., Hartman N.S., Whetzel D.L., Grubb W.L.* Situational Judgment Tests, Response Instructions, and Validity: a Meta-analysis // *Personnel Psychology*. 2007. Vol. 60. Is. 1. P. 63–91.
- Picard R.R., Cook R.D.* Cross-Validation of Regression models // *Journal of the American Statistical Association*. 1984. Vol. 79. Is. 387. P. 575–583.
- Schmidt F.L., Hunter J.E.* The Validity and Utility of Selection Methods in Personnel Psychology: Practical and theoretical implications of 85 years of research findings // *Psychological Bulletin*. 1998. Vol. 124. Is. 2. P. 262–274.
- Schippmann J.S., Prien E.P., Katz J.A.* Reliability and Validity of In-Basket Performance Measures // *Personnel Psychology*. 1990. Vol. 43. Is. 4. P. 837–859.
- Shadish W.R., Cook T.D., Campbell D.T.* Experimental and Quasi-experimental Designs for Generalized Causal Inference. Boston, MA: Houghton-Mifflin, 2002.

- Smith M.* Calculating the Sterling Value of Selection // *Guidance and Assessment Review*. 1988. Vol. 4. Is. 1. P. 6–8.
- Tabachnick B.G., Fidell L.S.* Using Multivariate Statistics. Fifth Edition. Pearson International Edition, 2007.
- Thornton III G.C., Rupp D.E.* Assessment centers in human resource management: strategies for prediction, diagnosis and development. Lawrence Erlbaum Associates, 2006.
- Wagner R.K., Sternberg R.J.* Tacit Knowledge: Its Uses in Identifying, Assessing, and Developing Managerial Talent // *Applying Psychology in Business: The Manager's Handbook* / Ed. by J. Jones, B. Steffy, D. Bray. New York: Human Sciences Press, 1991. P. 333–344.