



Отчет по валидации базового теста SkillCode на оценку навыков Soft Skills

1. Введение

Валидность — это ответ на вопрос: оценивает ли тест то, что мы хотим оценить, и насколько хорошо? Для этого первоначально нужно зафиксировать, какие параметры и для чего необходимо определять. Данный вопрос рассмотрен в разделе **Конструктивная валидность**.

В методике присутствуют блоки тестирования как интеллектуальных, так и личностных, поведенческих, коммуникативных, ролевых, мотивационных и других характеристик. Это означает, что эффективность методики в одной части блоков определяется оценкой правильности выполненных заданий, а в другой части оценкой соответствия результатов тестирования результатам, полученным при:

- интервью с экспертами-психологами с оценкой значимых проявлений вербальных и невербальных параметров;
- тестировании испытуемых другими известными валидными методиками: 16-ти факторный опросник Кеттелла, индикатор Майерс-Бриггс (МВТИ) с сопоставлением корреляции результатов в части идентичных конструкторов;
- проведении фокус-групп для оценки групповых взаимодействий респондента;
- моделировании ситуаций для оценки поведенческих реакций респондента;
- моделировании кейсов для проработки паттернов взаимодействия испытуемых в различных ролевых моделях, например в качестве руководителя и подчиненного;
- других практических методик оффлайн экспертной оценки.

2. Конструктивная валидность

Это сложное философско-методологическое понятие, которое введено относительно недавно – в конце XX века.

Конструкт – это некая разработанная психологами в результате наблюдений и аналитики сущность, отражающая те или иные свойства психики и поведения. Конструкты строятся из набора факторов, определяющих различные характеристики. В представленном тесте оценивается 36 факторов, которые условно отнесены к трем основным свойствам тестируемого:

- Коммуникация (состоит из 12 факторов)
- Мотивация (состоит из 11 факторов)
- Управление (состоит из 13 факторов)

Первоначальные опросники и задания формировались на основании и с учетом известных методик, современных научных исследований и методических материалов, которые оценивают выбранные факторы. В первой итерации экспертной группой были отобраны 45 факторов (по 15 в каждой группе), сформированы задания и опросники для их оценки после чего был проведен первый этап тестирований с последующим кросс-факторным анализом и предварительной оценкой конструктивной валидности. Работы по первому этапу проводились в общей сложности на выборке из 157 тестируемых, которая была сформирована в соответствии с генеральной совокупностью потенциальных испытуемых из: студентов последнего курса высших учебных заведений (20% тестируемых), офисных работников не

руководящих должностей и специалистов с опытом работы в занимаемой должности не менее 1,5 лет (30% тестируемых), руководителей среднего уровня управления с опытом работы в занимаемой должности не менее 1 года (40% тестируемых), руководителей высшего уровня компаний с количеством сотрудников не менее 100 человек (10% тестируемых), соотношение по гендерному признаку 59% мужчин, 41% женщин.

По результатам проведения 1 этапа валидизации были получены следующие результаты:

1. При рассмотрении валидности заданий с наличием правильного ответа, оценивающих интеллектуальные характеристики, определено распределение сложности вопросов.

Сложность	%% правильных ответов	%% заданий
Высокая	0% - 20%	16%
Выше средней	21%-40%	18%
Средняя	41%-60%	24%
Ниже средней	61% - 80%	21%
Низкая	81%-100%	21%

Вопросы распределены по сложности равномерно, около 20% на каждый уровень.

Дальнейший анализ показал, что вопросы с 0 правильных ответов (2) содержали недоработки, которые были исправлены.

Вопросы со 100% правильных ответов (3) были исключены как невалидные.

2. Определение шкал теста. Выполнялось в 3 этапа:

- Факторный анализ кросс-корреляций заданий. В результате были выделены кластеры заданий, коррелирующих между собой на уровне свыше +0,8 при статистической значимости 0,05.

- Содержательный анализ заданий: определение смысла и содержания заданий, входящих в кластеры.

- Компоновка шкал теста. На этом этапе было произведено выравнивание шкал: группировка по смыслу и внутренней связности в шкалы.

Всего было выделено 40 кластеров с высокой внутренней взаимосвязью. (Корреляция кластеров между собой – слабая положительная: от +0,3 до +0,5. Это означает, что шкалы действительно измеряют разные способности респондентов и могут проявляться в разных сочетаниях. 5 факторов были объединены с другими коррелирующими факторами.

3. После проработки результатов посредством проведения индивидуальных интервью, фокус групп, ситуационных и ролевых симуляций было получено соответствие результатов от 70% до 80% по 38 факторам и результаты валидности от 26% до 33% по 2 факторам. Данные 2 фактора с низкой валидностью были исключены из методики.

3. Для последующей оценки были выбраны 50 респондентов, которые были протестированы по ряду существующих валидизированных методик. По итогу сравнения результатов данных респондентов по разрабатываемому тесту с сопоставимыми однородными или схожими конструктами из существующих валидизированных тестовых методик были получены проценты соответствия результатов от 69% до 77% по 36 факторам и от 43 до 55% по 2 факторам. Полученное снижение результатов валидности относительно предыдущего этапа оценки можно объяснить неполностью однородным соответствием некоторых конструктов используемых существующих методик с разрабатываемой методикой, а также выделение 2 факторов с общей сниженной валидностью в сравнении с другими методиками.

Выводы:

Проведённый предварительный анализ показал, что тест по 36 факторам обладает достаточной валидностью, результаты респондентов соответствуют их должностному уровню, вопросы с наличием правильного ответа равномерно распределены по степени

сложности. Факторы с высокой корреляцией (5) были объединены с сокращением количества заданий этих общих факторов, факторы с низкой конструктивной валидностью и низкой сопоставимостью с другими методиками (4) были исключены из работы. Также были исключены отдельные задания (3) с низкой валидностью и внесены незначительные корректировки в некоторые формулировки без изменения смысловой сути заданий.

После завершения 1 этапа валидизации была проведена работа по масштабному тестированию на большом количестве тестируемых. Для этой задачи были привлечены респонденты, в соотношении, соответствующем структуре первоначальной выборки. Выборку представили 571 человек, в которую вошли выпускники ВУЗов, линейные сотрудники, менеджмент и руководители компаний из следующих сфер деятельности: производство, банковский сектор, агентства недвижимости, консалтинговые агентства, нефтегазовая отрасль, продажи. На втором этапе валидизации было продолжено в том числе определение критериальной валидности.

3. Критериальная валидность

Критериальная валидность тестовой методики демонстрирует, насколько тест отвечает заданным критериям. Среди наиболее важных критериев выделяют следующие:

1. **Критерий дискриминативности (различающей способности) заданий, имеющих правильный вариант ответа.** Задания не должны быть ни слишком простыми (которые решает 100%), ни слишком сложными (которые не решает никто). Следует установить и убрать, либо изменить задания с выполнением свыше 95% и ниже 5%. Эта проверка для заданий с наличием правильного ответа была сделана на 1 этапе валидизации. На втором этапе валидизации на большей выборке были получены следующие данные:

Сложность	%% правильных ответов	%% заданий
Высокая	0% - 20%	17%
Выше средней	21%-40%	19%
Средняя	41%-60%	25%
Ниже средней	61% - 80%	24%
Низкая	81%-100%	15%

Данные значения подтверждают полученные ранее результаты по равномерности распределения сложности заданий.

2. **Критерий нормальности распределения (несмещённости выборки).** Результаты теста по выборке должны соответствовать распределению Гаусса. При невыполнении данного требования окажется, что для испытуемых тест в целом слишком сложен, либо слишком прост (для заданий с наличием правильного варианта ответа). На нормальность распределения по ряду психологических факторов влияют параметры выборки такие как уровень должности и характер выполняемой работы, а также сфера деятельности компании. Учитывая, что значительная часть выборки представлена руководящим персоналом, то по некоторым факторам наблюдается смещение выборки в сторону более высоких значений, что является нормальным для данной выборки явлением. Распределение результатов по каждому фактору по результатам 2 этапа валидации приведен в Приложении №1.
3. **Критерий валидности по объективным факторам.** Значения по данному критерию были определены на первом этапе валидизации при проведении очных интервью с психологом, фокус групп, ситуационных и ролевых проработок. На данном этапе было получено подтверждение валидности выбранных конструктов и критериев валидности по объективным факторам с учетом корреляции с другими тестовыми методиками. В

рамках второго этапа валидизации при оценке критерия валидности по объективным факторам посредством комплексной оценки соответствия результатов теста по процедуре «ассесмент центр», в том числе с валидными методиками, на всех респондентах были получены значения соответствия результатов от 71% до 76%, что повторило результаты первого этапа валидизации.

Надёжность

Надёжность теста показывает, насколько его результат зависит от условий применения к одному и тому же испытуемому. Например, получим ли мы одинаковые результаты, если человек только проснулся или бодрствует, сыт или голоден, трезв или пьян, и т.д. и т.п. Или, насколько результаты теста могут меняться со временем. Для снижения внешних факторов на результаты тестирования перед началом его прохождения дается инструкция, в которой говорится о том, что необходимо располагать достаточным количеством времени, и находится в спокойном эмоциональном состоянии без признаков усталости.

При этом все личностные тесты очень зависимы от социальных условий. Например, в ММРІ есть утверждение: «Вы религиозны». Если представить, как воспримет это американец, или русский в 70-х, 90-х и 2020-х годах, то ответ будет иметь абсолютно разные смыслы, вплоть до противоположных. Нередки случаи, когда тестируемый может показать абсолютно разные результаты с небольшим интервалом между тестированиями. Например, был случай, когда по тесту ММРІ с интервалом в 2 недели человек выдал абсолютно разные результаты, как если бы его проходили 2 разных человека. Оказалось, что в этот период у человека умер близкий родственник.

Надёжность измеряется корреляцией результатов тест-ретестовых проб, например, коэффициентом Пирсона. Тесты с высокой надёжностью выдают корреляцию результатов при уровне значимости 0,01 на 0,8-0,9. На практике корреляция 0,7 считается приемлемой для выборки от 100-150 испытуемых. При этом на надёжность сильно влияет выборка испытуемых. Это означает, что тест можно считать надёжным только для определённой, однородной относительно измеряемого качества, группы испытуемых.

При проведении точечных тест-ретестовых проб среднее значение корреляции результатов тестирования разработанной методики составило 0,73.

Приложение №1

Анализ нормальности распределения результатов тестирования по 36 факторам



