

# Validation of Shadowmatch™

**Shadowmatch Introduced:** Shadowmatch is a 'Black-box' type tool that presents an individual with a list of tasks in order to determine habits in the behaviour of the individual. It simulates tasks for the individual to indicate how he/she will act by selecting from a list of multiple answers. The system then identifies trends in the way an individual has indicated how he/she will act in the specified circumstances and calculates the consistency with which answers were selected. The end result is a graph that indicates the level to which these habits are embedded in the behaviour of the individual as per a set of behaviour definitions. A high score indicates that the individual has consistently selected answers that indicate a strong preference towards behaving in a specific manner. A low score indicates that the individual didn't consistently select answers that would represent congruent behavioural patterns of the specific nature calculated as a habit. On the list of 19 habits, the score can even be less than zero. This indicates an anti habit also referred to as a counter habit. The person then indicated a habit against the habit being calculated.

## Statement of Experiment:

**Assertion:** Shadowmatch asserts that by having an individual complete a list of tasks (in the format of a questionnaire based worksheet) Shadowmatch can discriminate the level or degree to which habits are embedded in the behaviour of an individual as per a set of defined behaviour categories.

The requirement is to design an experiment by means of which the above assertion can be tested with a view towards validation. Our experiment consists of the following:

1. Inviting eight 'experts' to complete a total of 36 Shadowmatch worksheets profiling 18 separate behaviour categories such that 18 result in a score of 70 or more (demonstrating the presence of a habit imbedded in the behaviour of an individual) and such that 18 result in a score of 10 or less (demonstrating the absence of a habit embedded in the behaviour of the individual).
2. The eight experts were not told in advance which behaviour pattern they would profile. Once seated, each was provided with access codes to complete a Shadowmatch worksheet and each was randomly assigned a separate behaviour category with a definition of that behaviour category.
3. For each behaviour category, an expert was tasked to complete two Shadowmatch worksheets.

- a. One: complete a worksheet, consistently selecting for each question the answer that they (the tasked expert) felt best described behaviour that was consistent with the behaviour definition supplied, and,
- b. Two: complete another worksheet, consistently selecting for each question the answer they felt was most anti / contra the behaviour as per the behaviour definition.

## Interpretation and Its Reasoned Basis

In considering the appropriateness of the above experimental design, it is needful to forecast all possible results of the experiment and to have decided without ambiguity what interpretation shall be placed upon each result. In this instance, we set out the possibilities as per the mathematics of permutations and combinations for the 18 results out of 36 total tests that targeted a score of 70 or more. In determining the full set of possible permutations, it is noted the order does not matter and a particular result can only be used once. Consequently, the number of possible combinations is the binomial coefficient:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

where for e.g.  $3! = 3 \times 2 \times 1 = 6$ ; [ ! stands for factorial ].

For our calculation,  $n = 36$  and  $k = 18$ . There are a total of 9,075,135,300 (just over 9 billion) possible sets of answers ranging from all 18 targeted behaviour categories scoring 70 or more to all 18 targeted behaviour categories scoring less than 70. At best, the experts can complete the worksheets so consistently as per the behaviour definitions that all 18 targeted behaviour categories result in a score of 70 or more. In terms of probability, experts without any faculty of discrimination against the definitions would emulate this result (all 18 scoring 70 or more) in one trial out of 9,075,135,300. The complete set of probabilities is enumerated in the table 1 following:

**Table 1: Full set of possible results (normal distribution curve):**

Behaviours Scoring 70+ out of 18	Behaviours Scoring less than 70	Frequency	Significance	Confidence
18	0	1	0.00000001%	99.99999999%
17	1	324	0.00000357%	99.99999643%
16	2	23,409	0.00025795%	99.99974205%
15	3	665,856	0.00733715%	99.99266285%
14	4	9,363,600	0.10317863%	99.89682137%
13	5	73,410,624	0.80892044%	99.19107956%
12	6	344,622,096	3.79743204%	96.20256796%
11	7	1,012,766,976	11.15980030%	88.84019970%
10	8	1,914,762,564	21.09899743%	78.90100257%

9	9	2,363,904,400	26.04814498%	73.95185502%
8	10	1,914,762,564	21.09899743%	78.90100257%
7	11	1,012,766,976	11.15980030%	88.84019970%
6	12	344,622,096	3.79743204%	96.20256796%
5	13	73,410,624	0.80892044%	99.19107956%
4	14	9,363,600	0.10317863%	99.89682137%
3	15	665,856	0.00733715%	99.99266285%
2	16	23,409	0.00025795%	99.99974205%
1	17	324	0.00000357%	99.99999643%
0	18	1	0.00000001%	99.99999999%

## The Test of Significance

In relation to test of significance, we can conclude that something is demonstrated experimentally when we can conduct an experiment in such a way that it will rarely fail to give us a result that can be interpreted against a statistical significance score. Table 1 above details all the possible results of the experiment as pertains to demonstrating a habit embedded in the behaviour of an individual (that is a score of 70 or more resulted for the behaviour category targeted). It is noted that the sum of the frequency above equals the total number of possible combinations of 9,075,135,300. The significance score is calculated as: [ frequency divided by 9,075,135,300 ]. In deciding the level of significance to use, we want to avoid high degrees of success that can result from mere chance. Hence we target a level of significance of 0.01 for a 99% plus confidence index in our results.

## The Null Hypothesis

Our exploration of the full set of possible results of the experiment has resulted in a statistical test of significance by which the actual results can be divided into two classes of opposed interpretation. The two classes of results which can be distinguished by our test of significance are one (forthwith referenced as the 'null hypothesis'): *the hypothesis that the experts' collective efforts to discriminate the presence of a strong habit as well as the absence of a habit embedded in an individual's behaviour as per a set of behaviour definitions are not enabled by the Shadowmatch worksheet,*

and two (forthwith referenced as the alternative hypothesis): *the hypothesis that the experts' collective efforts to discriminate the presence of habits (strong and absent) embedded in an individual's behaviour as per behaviour definitions are enabled by the Shadowmatch worksheet.* It is noted that this experiment ('validation of Shadowmatch') has been created to give the evidence a chance to disprove the null hypothesis and not to prove or establish it.

It is tempting to argue that if the validation of Shadowmatch (the experiment) can disprove the null hypothesis, we must consequently be able to prove the opposite hypothesis namely that

experts can discriminate the presence of habits embedded in an individual's behaviour as per a set of definitions using the Shadowmatch worksheet. However, this opposite hypothesis is ineligible as a 'null hypothesis' because it is inexact. If we would add the words '**always** be able to discriminate the presence of a habit embedded', we would then have an exact hypothesis that can operate as a 'null hypothesis'. But it is easy to see that this hypothesis can be disproved by a single failure whilst never being proved by any finite number of experiments. From this example, it is clear that the 'null hypothesis' must be precise and unambiguous in order for it to supply the basis of the 'problem of distribution', for which the 'test of significance' is the solution.

### **Randomisation: The Physical Basis of the Validity of Shadowmatch**

We have discussed that the validation of Shadowmatch involves testing the null hypothesis that the Shadowmatch worksheet does not enable experts to discriminate the presence of an embedded habit. We have also assigned, as appropriate to this hypothesis, a frequency distribution of occurrences, based on the equal frequency of the 9,075,135,300 possible ways of assigning the results of 36 Shadowmatch worksheets to two habit sets (embedded and absent) of 18 behaviour areas each. This is in fact the frequency distribution for a classification by pure chance.

We have now to consider the physical conditions of the experimental design required to substantiate the assumption that, if the link between experts being able to use the Shadowmatch worksheet and behaviour definitions to model the degree of habits embedded in an individual's behaviour is absent, the results of the experiment conducted will be completely controlled by the laws of chance. It is easy to see how the physical conditions of the experiment can deliver results not completely controlled by the laws of chance. For example, if, for each question in the Shadowmatch worksheet, an expert had a choice between two answers highlighted in different colours to consistently indicate the absence or presence of a habit embedded in the behaviour set under modelling, then the obvious difference in the format of the question answers would ensure that a set of both high scores and low scores result. This would happen in 50% of all trials and the sets would either be all right or all wrong and the test of significance would be meaningless.

In this particular instance, we needed to take the precaution of randomisation to guarantee the validity of the test of significance against which to decide the experiment results. This was done through behaviour sets being randomly assigned to each of the eight experts. For each specific behaviour area the expert was required to complete a Shadowmatch worksheet to model both the presence and absence of a habit embedded in an individual's behaviour as per the behaviour definition. The worksheet targeting a specific behaviour could only be completed once by the expert in question. Second attempts at modelling the behaviour are not a part of the experiment results – first attempts only have been included.

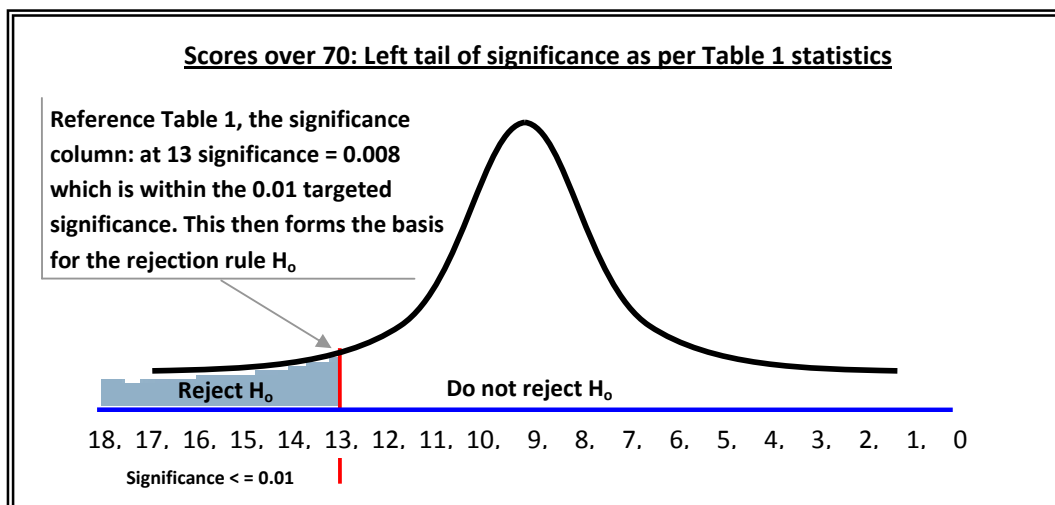
With respect to any additional refinements that can be made to increase the sensitivity of the experiment, we now have in place the description of a valid experiment as well as the test of significance by which to determine its result. It remains for us to translate the hypotheses into numerical constructs, to present the actual results of the experiment and to conclude on the validity of Shadowmatch.

### Hypotheses $H_0$ and $H_a$ and Rejection Rule $H_0$

Null Hypothesis:  $H_0$ : Shadowmatch does not discriminate the presence of habits embedded in an individual's behaviour.

Alternate Hypothesis:  $H_a$ : Shadowmatch does discriminate the presence of habits embedded in an individual's behaviour.

We want a 99% chance of discriminating the presence of habits embedded in an individual's behaviour. Hence, the level of significance for the hypothesis test is 0.01 (or 1%). If you reference Table 1 earlier, you will see that the significance scores for a total of 13 to 18 tests scoring greater than or equal to 70 are all less than 0.01 (1%) and thus within the 0.01 significance level. This band thus forms the basis for our rejection rule  $H_0$  detailed below.



Given Table 1, a significance level of 0.01, the above visual and the fact that the same logic applies to the results targeting scores of 10 or less, we can establish the following rejection rule:

Reject  $H_0$ : If 13 or more of the 18 worksheets completed to show the presence of a strong habit embedded in an individual's behaviour score greater than or equal to 70.

OR

If 13 or more of the 18 worksheets completed to show the absence of a habit embedded in an individual's behaviour score less than or equal to 10.

## The Experiment Results

The expert group of eight could successfully replicate the presence /absence of behavioural habits as is presented in table 2 below. (Manipulating the results equal to or above 70% should be seen as 100% successful and creating a result of less than or equal to 10% in order to anticipate the absence of a behavioural habit, that should also be seen as 100% successful).

**Table 2: Experiment Results (36 results made up of two sets of 18)**

No	Habit.	Forced High	Count of Successful	Forced Low	Count of Successful
1	Propensity to own	71%	1	19%	
2	Propensity to hand-off	64%		13%	
3	To Simplify	70%	2	9%	1
4	Resilience	83%	3	-23%	2
5	Propensity to Change	70%	4	3%	3
6	Frustration Handling	73%	5	-6%	4
7	Team Inclination	96%	6	4%	5
8	Individual Inclination	72%	7	6%	6
9	Self Motivation	81%	8	-2%	7
10	Routine	55%		12%	
11	Problem Solving	79%	9	-6%	8
12	Responsiveness	88%	10	3%	9
13	Innovation	83%	11	-14%	10
14	People Positive	85%	12	-14%	11
15	Discipline	69%		12%	
16	Conflict Handling	67%		3%	12
17	Altruism	80%	13	-10%	13
18	Self Confidence	78%	14	-9%	14

## Conclusion

Set one produced 14 scores of 70 or more and set two also produced 14 scores of 10 or less. As this is greater than the 13 limit as specified in rejection rule  $H_0$ , we **reject** the null hypothesis that Shadowmatch does not discriminate the presence of habits embedded in an individual's behaviour and accept the alternate hypothesis ( $H_a$ ) that Shadowmatch does discriminate the presence of habits embedded in an individual's behaviour with a significance of 0.01. This means that Shadowmatch gives a user of Shadowmatch a 99% probability of discriminating the

presence of habits embedded in an individual's behaviour as per Shadowmatch's behaviour definitions.

## **Participants**

1. Guy Krige: Independent Business Consultant
2. Erna Gerryts: Independent Control Group Participant
3. Theo Bezuidenhout: Independent Sport and Counselling Psychologist
4. Madi du Toit: Deloitte
5. Nelius Volschenk: Deloitte
6. Pieter de Villiers: Shadowmatch Representative
7. Lizette Bester: Shadowmatch Representative
8. Hestie Byles: Psychologist University of Pretoria (partial participant)

## Appendix A: Habits Interpreted against a Point Scale

The calculator in Shadowmatch is such that the relative score can vary from the extreme low – even a minus score in the extreme instances – to a very high score. The following framework of points is an indication of what a low and a high score will be:

Less than 10 points is a very low result.

Between 10 points and 30 points is a low result but not radically low.

Between 30 and 50 points indicates a non radical but strong result.

Between 50 and 75 points is a strong to radically strong result.

More than 75 points is a radically strong result.

Although the purpose of this validation is not to interpret these results, it is important to indicate how these points are interpreted as a relative indicator of the behavioural habits and the level to which the specific habit are defined. The example will be Problem Solving as a habit. Shadowmatch defines Problem Solving as follows:

**Problem Solving:** This is the habit of engaging with challenges on a conceptual, social and practical level and successfully managing these difficulties/challenges towards resolving them. People with a strong embedded habit of problem solving easily becomes intrigued by challenges and riddles to be resolved.

The results on problem solving will indicate the following problem solving behavioural preferences:

**Less than 10 points:** This result will indicate that the individual has a low inclination towards problem solving. He/she will rather find alternative ways of dealing with the problem, this might vary from handing it to somebody else to resolve, ignoring it, solving it when he/she is part of a team or even living with the problem without ever resolving it. The individual will only engage with problems under very specific circumstances.

**Between 10 and 30 points:** This indicates that the individual will engage with a problem, not as a personal adventure and a choice, but when necessary. It might even be that the individual will engage with the problem if it is easy to resolve or if it presents something of special interest or even if it is related to a unique skill that the person has.

**Between 30 and 50 points:** This individual will engage with problems with relative ease. He/she will be selective in the type of problems he/she prefers to resolve but for the majority of day to day problems the individual will engage with the problem with relative comfort.

**Between 50 and 75 points:** This individual – for whatever reason – finds meaning in tackling and resolving problems. Although he/she might still be selective in the type of problem he/she prefers to resolve, he/she will engage with the majority of problems in life with the purpose of resolving them. These individuals might even find some form of personal fulfilment in solving problems. It is a habit for them to engage with challenges in order to get a solution.

**More than 75 points:** These individuals find it difficult to walk away from any problem. In the extreme they tend to become so involved with a problem that it might even influence their normal day to day lives. They can become hooked on a problem and engage with it over long periods of time in order to resolve it. They have a very prominent habit towards engaging with problems, they might even feel frustrated if they are faced with a problem that cannot be resolved. In the extreme, these people might become so intrigued with problems and problem solving that they will start a hobby that entails problem solving.

All the habits tracked by Shadowmatch are more or less in the same weighting framework as indicated above with regards to Problem Solving. The cut-off points are not absolute, they are soft borders. It must also be emphasised that Shadowmatch doesn't measure the way people think, their feelings, temperaments, tastes, moral character, beliefs, expectations, desires, values, or their personalities. It identifies behavioural patterns or habits that could best be described as follow:

*A **habit** is an action that repeats itself with no (or minimal) conscious planning. It repeats itself when the situation is conducive to such behaviour and the person has a goal of fulfilling his/her need in some way by doing what he/she normally does. This action can then become a habit.*

## Appendix B: When is Behaviour a Habit?

Any behavioural pattern can become a habit. When is behaviour a habit? What else do we need to know about habits?

- Any behaviour that repeats with minimal/no planning can be a habit.
- Habits are learnt behaviour.
- Habits are associated with some meaningful goal/purpose.
- Habits are formed through repetition.
- Habits are very predictable.
- Any behaviour can become a habit. (Getting up early, solving problems, working alone, disengaging when frustrated, etc).
- Habits are extremely difficult to break.
- When an individual cannot freely live by his/her habits, he/she finds it extremely difficult to function properly.
- People living their lives along well established habits show highly predictable behaviour. The opposite is also true.
- All habits are not equally well established.
- Habits can transcend contextual boundaries. (A habit developed in one context can manifest in another context without a clear link to the environment where the habit has been developed).
- Habits are always attached to some kind of meaningful purpose for the individual. This purpose might be something like avoiding discomfort, being successful, being efficient, etc.
- Habits can form pairs or tandem habits. This happens when two or more behavioural types that are related to each other are well developed. E.g. People Positive Behaviour and Altruism. When these two habits are strong, they contribute to each other and the behaviour becomes more prominent.