

**Knowing yourself as a change agent:  
A validated test based on a colorful theory of change**

**Léon de Caluwé, Hans Vermaak**

**Draft article under review**

**ABSTRACT**

The article presents a questionnaire that measures individual change preferences based on a meta-theory of five paradigms of change. We describe its construction and improvement over a 13-year period, during which time more than 100,000 people have used the test as an instrument for self-reflection. Our analysis shows how change preferences are distributed among different demographics. It also reveals that change paradigms are differentiated along two dimensions. We hypothesize that one dimension corresponds to the type of change leadership (top-down versus bottom-up) and the other to the type of change relationship (subject-object versus subject-subject). The questionnaire has proven effective in sparking discussions that aid professionalization and collaboration. We discuss how the test can be used most effectively.

## **Knowing yourself as a change agent: A validated test based on a colorful theory of change**

In the last two decades there has been a rise in publications that advocate a multi-paradigmatic view of organizational change (e.g., Van de Ven & Poole, 1995; Beer & Nohria, 2000). This trend towards pluralism is good news given the diversity of organizational issues that cannot be dealt with effectively with a uniform approach. Change agents need to be aware not only of the range of approaches, but also of their own preferences, capabilities, credibility, and limitations in terms of this range. This implies a need for reflective practice (e.g., Schön, 1987). In this article we discuss an instrument that can aid such reflection: a style test for change agents that creates a profile of their sympathies and antipathies for contrasting change approaches. We have based this test on a meta-model of change we started developing almost 20 years ago: the color model. The test has been freely available for the last 15 years and has been used by more than 100,000 people. During this time, we have continued refining the test to increase its validity. In this article, we describe this process and share the results of the latest version of the test utilized by a population of about 3,500 people. Lastly, we discuss how the test can be most effectively used and how the results can be interpreted.

### **THE COLOR MODEL**

The color model distinguishes between five fundamentally different ways of thinking about change, with each color representing a paradigm of different beliefs and values about change. Each of these paradigms is labeled with a color, only as a kind of “shorthand” and without much symbolic connotation, and each represents different traditions or schools of thought in our field. The colors have their own characteristics in terms of type of interventions, diagnostic models, roles, and outcomes. Together they comprise a meta-theory of change that has several applications, one of which we focus on here: to reflect on one’s own preferences and possibilities as change agent. A more comprehensive description of the color model (De Caluwé & Vermaak, 2004) and its development and manifestations (De Caluwé & Vermaak, 2015) are available. Table 1 summarizes the core components, underlying assumptions, and key traits of each of the five colors (Authors, xxxx).

-----  
Insert Table 1 about here  
-----

The color model can be a lens through which to look at one’s background, competencies, portfolio of assignments, image and credibility, networks, and so on. We like to point out

that the colors refer to belief systems and deeply held assumptions about the nature of change, which implies that they may not always be consciously chosen to fit the issue at hand. Our belief systems can cause us to be attached to certain preferences, which show up not only in terms of what we think, say, and do, but also are part of how we perceive ourselves. Our style of working, the values we espouse, and the traditions in which we take part can become part of our (professional) identity. They may cause us to have strong antipathies or “allergies” to other colors on the spectrum. We have often noticed that people are not fully aware of their preferences and this can have a negative impact in terms of not knowing one’s limits, not respecting other points of view, or not exploring different strategies when need be. In such cases, feedback from others can be of help, like a mirror. The more such feedback is gathered from different sources, the more reliable such a mirror will be. We developed a test to assist in such self-reflection: a questionnaire to measure preferences about change. Based on their answers to the questionnaire, people can identify their own dominant beliefs.

## THE BASIC CONSTRUCTION OF THE TEST

### **Forced choice**

We decided early on to construct a test based on forced choice, as it seemed to offer a good compromise between ease of use and reliability of measurement. At the outset, we tried three different types of tests. The first one was a Likert five-point scale (Test 1 in table 2) with sixty items to be answered (agree/neutral/disagree). We quickly moved to an ipsative-style test with thirty items (Tests 2–4 in table 2), based on a simple forced choice between two alternatives (A/B). The test that we have used predominantly since 2000 is a test with ten to twelve items (Tests 5–18 in table 2) based on a more subtle forced choice in which participants have to distribute points between five alternatives: a “test of points” (“*puntentes*” in Dutch). The combination of reducing the number of items and increasing the subtlety of forced choice allowed us to get results that were still reliable but offered more ease of use. The test now allows respondents a) to fill out the questionnaire in a short amount of time (10–15 minutes), b) to create their own profile without external help, and c) to get their results immediately either by adding their own scores on paper or having them calculated online. It also allows substantial data to be gathered with ease, facilitating its use in research, teaching, and other group settings.

Another advantage of forced choice is that it nudges people to show their “true colors” and makes it harder to give “middle of the road” answers. It forces respondents to discern the values and beliefs they hold most dear. The use of closed questions allows them to do so based on their “gut feeling,” and without prior knowledge of the model behind the test. One disadvantage of this method is that respondents cannot give nuanced or tailored answers; because of this, relevant data may get lost and respondents may get frustrated

because they can't choose the answer that is most true for them (Van der Velde et al., 2008).

### **Example of a test item**

In order to provide a sense of the test, we include here an example of one of the twelve questions that is part of the final "test of points" (see the links at the end of the article)

In my opinion change can only be successful if:

- a. It is supported by the most important managers.
- b. The employees support the change.
- c. Clear objectives have been set beforehand.
- d. Employees gain new insights.
- e. The strengths and energy of those involved are activated.

The respondents are asked to distribute eight points over these five alternatives, which makes it hard to distribute the points evenly. They are instructed to distribute the points based on how well the statements match their convictions. They can, for instance, give eight points to one choice, four points each to two choices, or give one, three, and four points to three choices. After having distributed all of the points, the respondent's score can be interpreted. In the above example, letter a) refers to yellow, b) to red, c) to blue, d) to green, and e) to white. The points for each question are added up by color and provide an overall profile of one's preferences (high scores) and antipathies (low scores).

### **The content of the test: The items**

The questions inquire into many aspects of change in order to create a color profile. Some relate to underlying assumptions directly, others indirectly. The diversity of questions makes the test more reliable. The division of the items is as follows:

- One item relates primarily to how people change (item 11).
- One item relates primarily to how organizations change (item 5).
- Four items relate primarily to characteristics of change processes, such as key activities or interactions (items 2, 4, 9, 10).
- Three items relate primarily to the context of a change process, such as conditions, measure of success, or values (items 1, 7, 8).
- Two items relate primarily to characteristics of change agents, such as their role or competences (items 3 and 6).
- One item relates to a resonance with proverbs that capture the belief systems of a color (item 12).

The statements within each item are derived directly from the color theory itself. Part of the process of refining the test was to create statements that were formulated in a way that

did not paint one color in a more positive light than others, which meant we needed to correct our own biases in describing the color model. We learned first-hand of the problem of incommensurability of meta-models: there is no objective way to talk about belief systems (e.g., Scherer & Dowling, 1995). It took us a decade to minimize such biases.

## **THE INCREMENTAL DEVELOPMENT OF THE TEST**

Table 2 gives an overview of the different tests and samples used to refine the questionnaire and its interpretation. The first column shows that 18 samples were used between 2000 and 2013. The second column in table 2 shows the type and version of each test: there is one version of the first type of test (Test 1), three versions of the second type of test (Tests 2–4), and five versions of the third type of test (Tests 5–18). This third type of test is the “test of points” that we have used and researched the most.

The test of points was refined four times based on a statistical analysis of the number and distribution of items and a rewording of the statements within the items. These improvements were researched and documented in collaboration with master’s degree students. The first improvements to versions 2 and 3 were based on Oort (2006) who analyzed almost 2,700 questionnaires (Test 6 in table 2). The next improvement to version 4 was based on Lankreijer’s (2007) analysis of 280 questionnaires (Test 7 in table 2). Tummers (2009) validated this version of the test with over 1,700 respondents, and found clear correlations between the statements and the colors they are supposed to represent (Test 9 in table 2). The last improvements to the test were made on the basis of Tummers’s work. Pietersen (2013) used the fourth and the fifth (final) versions of the test for his analysis with a total of almost 3,500 respondents (Tests 14–16 in table 2).

Because of the ipsative character of the data, a factor analysis was regarded as unsuitable (see Blinkhorn et al., 1988). Ipsative data typically produce bipolar factors: this is caused by the forced choice format, where choosing one option inevitably means not choosing the other. However, in real life, if you have to choose between fish and meat, and you choose meat, it does not mean that you do not like fish. In general the correlations of ipsative data are negative and lower than the correlations of normative results (see, e.g., Loo, 1999). For this reason, Tummers (2009) conducted a multidimensional scaling analysis (MDS). MDS visualizes the distance between variables: items that are perceived to be similar will fall close together on a perceptual map, and items that are perceived to be dissimilar will be further apart (e.g., Cooper and Schindler, 2008). Tummers’s MDS analysis showed that the distance between the test answers corresponding to one color tended to be shorter than the distance between the test answers corresponding to different colors. This proves that a common factor (a color) underlies test answers.

It is tempting to discuss the data collected through online versions of the test, given their large samples of more than 80,000 respondents (Tests 8 and 12 in table 2). We choose not to for two reasons. The first is that the online test is good for teaching, but the software is not geared for research: the dataset is condensed to simple management information that allows for little statistical analysis. The second reason is that we regard the online test as less reliable than paper tests because we have no information about the way the questioned are answered (such as the time or care spent on it). We observe, for instance, that more than 10% of online tests are filled out incompletely (these are not included in the data presented here). Our discussion instead focuses on the research done by Pietersen (2013), as he used the final version(s) of the test, had a large sample of respondents, collected the data under controlled conditions, and did the most robust statistical analysis.

-----  
Insert Table 2 about here  
-----

## MAIN RESEARCH OUTCOMES

Pietersen (2013) analyzed a sample of 3,995 questionnaires, collected between 2008 and 2012 at more than 150 seminars organized for people in leadership positions. All responses were recorded in situ: none of the data was submitted later or collected online. The total sample used here (after deletion of incomplete sets) is 3,687 (Tests 14–16 in table 2). It represents a cross-section of people in different leadership positions in organizations. More than 90% of the respondents have completed higher vocational education. More than 80% is part of the age category 35–44 or higher.

### **Reliability**

To measure internal consistency, a Cronbach alpha analysis was performed. This indicates how well the items in one set are positively correlated to one another for each factor (in our case, for each color). For Yellow that is:  $\alpha = .58$  (Test 14) and  $.58$  (Test 15); for Blue:  $\alpha = .76$  (Test 14) and  $.75$  (Test 15); for Red:  $\alpha = .52$  (Test 14) and  $.48$  (Test 15); for Green:  $.62$  (Test 14) and  $.67$  (Test 15); and for White  $.62$  (Test 14) and  $.58$  (Test 15). A desirable minimum is  $.60$  (Sekeran and Bougie, 2009), but ipsative data tend to show lower Cronbach's alphas than normative data (Saville and Wilson, 1991). In any case, the reliability cannot be increased by deleting answers from any of the 12 items.

### **Descriptive statistics**

The descriptive statistics are shown in table 3 for three samples of the latest two versions of the test. The average scores in both of Pietersen's (2013) samples are identical and differ little from those of Tummers's sample (2009). The standard deviation of the three samples is also similar. The standard deviation of Blue is the highest in both versions of

test; those of Red and Yellow are the lowest in both versions of the test, which means that for these two colors respondents' scores resemble each other most.

-----  
Insert Table 3 about here  
-----

### **Multidimensional scaling (MDS)**

The number of dimensions is identified by analysis of S-stress value. For the fourth version of the test (Test 14 in table 2), the value is between “good” and “excellent.” For the fifth version of the test (Test 15 in table 2), the value is between “fair” and “good.” Both versions appear to be based on two dimensions. Figure 1 shows all the items and related scores in the MDS analysis when they are aggregated by color and plotted. A similar pattern emerges for both tests. The only noticeable difference is that Red moves a little upwards in the latest test to more of a middle position.

-----  
Insert Figure 1 about here  
-----

Kruskal and Wish (1978) state that each factor should be clearly separated from others when they are plotted. This is definitely the case here: the colors occupy positions that are almost at optimal distance from one another. These positions had improved compared to earlier results, such as found by Oort (2006). The colors are now clearly differentiated from one another in the test.

### **What are the dimensions?**

We can distinguish two dimensions in the MDS plot. The horizontal dimension clearly separates Blue and Yellow from White and Green. The vertical dimension separates Yellow and Green from Blue and White. In both dimensions, Red takes up a middle position. The contrasts on the horizontal dimension are in line with previous research where we found a gap between Yellow and Blue on one side, Red in the middle, and Green and White on the other side (Van Nistelrooij et al, 2007; Knoop et al, 2009). We have not done previous research on the contrasts within the vertical dimension. The MDS analysis raises a question about how to conceptualize the dimensions. What would explain such a spacing of the colors? We offer the following explanation as a basis for further discussion.

We think the horizontal dimension (1) may refer to a preference for a type of change leadership. Negative scores correspond with a top down approach to change (Blue and Yellow), where agency is centralized. Blue-print change is generally designed and

executed by people who are trusted and mandated on the basis of their expertise. Yellow-print change is generally initiated and controlled by people on boards or in management who wield power through hierarchy. Both colors use a top-down approach based on underlying beliefs that change happens because of rational analysis, planning, and implementation (Blue), or because of a power coalition (Yellow).

In contrast, a positive score corresponds with a preference for a bottom-up approach to change (Green and White) or, to be more precise, an approach in which agency is decentered. Green-print change is driven by people's eagerness and ability to learn. Such a change may benefit from facilitation, but this is not meant to lessen participants' active stance. White-print change is often initiated and shaped by "tempered radicals" (Meyerson, 2003): people who care enough about an issue to take it up voluntarily. Both colors demonstrate an underlying belief that local ownership drives incremental change.

We think the vertical dimension (2) may refer to a preference for a type of change relationship. A positive score corresponds with subject-object relations, where a few people are active, knowledgeable, and influential, and others follow (e.g., Hosking, 2006). In such a view of change there can be no leaders without followers and vice versa. In Blue-print change, the experts and project managers do the analysis, planning, and directing of the change. Others follow their lead because they are put in charge, formally, for good reason, as they are "in the know." In White-change, the "tempered radicals" make sense of underlying dynamics, see new opportunities, and enroll others to take part in innovation. Here too, there are a few people in the lead because they are one step ahead of others, though not in a hierarchical sense. Leadership in both cases is not shared or distributed.

The opposite seems true for a negative score on this vertical dimension. This corresponds with subject-subject relations, where change is a collective endeavor (e.g., Kessener & Termeer, 2006). This is most pronounced in Green-print change, where learning is deemed something that happens through interaction with others. Regardless of whether learning happens by way of inquiry, experimentation, exercises, or teaching, meaning is created through conversations. In Yellow-print change, negotiations are the key to forming power coalitions and these too are created in interaction. Both approaches share the underlying notion that change is co-created with those involved and that many complementary contributions deepen the impact. Leadership is shared and people make sense of change together (e.g., Wierdsma, 2007).

Red-print change scores in the middle on both dimensions, which indicates an effort to somehow combine opposites. In the horizontal dimension, this refers to an attempt to reconcile centralized leadership with allowing the people involved some influence. It is an approach in which the direction of the change and its planning are still top-down, but



implemented with those involved. Thus the top-down approach is tempered to allow for participation, while trying to still maintain coherence and direction. In the vertical dimension, Red-print's in-between score indicates an effort to reconcile leadership by a few with the sense of community among the many. It is an approach with a clear division of roles and responsibilities that still tries to get as many people on board as possible.

The in-between position may be perceived as an ambivalent reaction to contrasting worlds. It is less controlled than Blue, less coercive than Yellow, less inquisitive than Green, and less entrepreneurial than White. It runs the risk of being a little bit of everything and not excelling in anything. Lewis (2000, p. 763) describes such ambivalence as “the compromise of conflicting emotions within lukewarm reactions that lose the vitality of extremes.” However, a Red-print approach can also try to reconcile opposites and explore transformative ways to deal with the paradoxical tensions between the colors. If it succeeds, it may be experienced as a process that somehow integrates contrasting values. We observe that such integration is, as yet, not all that common in terms of both ambition and realization. In a recent study, Smith and Lewis (2011) highlight organizational tensions, such as between learning (Green) and performing (Yellow), and describe how our actions can easily create self-fulfilling prophecies. The preference for an ambivalent solution could be only a concealing tactic, one that temporarily reduces discomfort yet eventually intensifies tensions and hampers performance. In contrast, a continued inquiry into divergent values and ways to interrelate them can produce recurring moments of transcendence and peak performance.

### **Differentiations between populations**

Are there correlations in terms of demographics? When we take the largest sample studied by Pietersen (2013) with a recent test of points (Test 14 in table 2), we observe some patterns. Pietersen analyzed a sample of about 2,700 people in terms of gender, age, employment sector, educational level, and leadership position. We share some results here to underscore that such correlations are often significant:

- In terms of gender, male respondents score significantly higher on Blue and Yellow than female respondents ( $p \leq 0.01$ ). In turn, female respondents score higher on Green and White than male respondents ( $p < .01$ ). There is no significant difference when it comes to Red.
- In terms of age, Blue scores decrease with age and White scores increase with age. While Yellow scores are low at younger and older ages, they peak among those in early career stages (aged 25–34). Red stays reasonably constant over time. For these four colors, the age correlations are significant ( $p \leq 0.01$ ). Green scores are not significantly correlated to age.
- There are correlations to be found between color preference and the type of sector in which people work. Pietersen (2013) compared and contrasted the scores for 13

different sectors. For instance, Yellow and Blue are most represented in accounting, auditing, and professional services, but least in education and consulting/interim management. Red is most represented in IT/ICT fields, and least in consulting/interim management. Green is most represented in education, and White is most represented in consulting/interim management. In contrast, Green and White are least represented in the fields of accounting and auditing. All these contrasts are significant ( $p < .01$ ).

- Respondents' preferences also correlate to education levels, distinguished as (a) secondary school, (b) vocational education, (c) higher vocational education, and (d) university education. Yellow appears most represented among university graduates and least among graduates of secondary school. Blue is most represented among graduates of vocational school and least by those of secondary school. Secondary school graduates score higher on Green and White than others, while Green is the least represented among university-level respondents and White is the least represented among vocational-school respondents. Such contrasts are significant for all colors other than Red ( $p < .01$ ).
- Lastly, leadership position is correlated to color preferences. The sample was analyzed for contrasts between managers, people in staff departments, and external change agents. Yellow and Blue are more represented in staff departments than the other two, Yellow least among managers, and Blue least among external change agents. Red is most represented among managers and least among external change agents. Green and White are most represented among external change agents and least in staff departments. These differences are significant for Yellow, Red, and White ( $p < .01$ ), and not as significant for Blue or Green ( $p = 0.5$  and  $.07$  respectively).

These correlations are open to interpretation. In fact, when we teach, we often engage with participants in discussion about how their background, type of work, and so on could be related to their change preferences. We think such hypothesizing is a useful learning exercise. We like to illustrate such hypothesizing here with the above findings with regard to the color preferences based on leadership positions. One may argue that it stands to reason that people in staff departments who have the least formal power, might want to wield influence. Given that Yellow and Blue are the colors that would dominate over other color strategies when it comes to conflict, it makes sense that these colors are their preferred choice. As managers have formal power, they have less need to wield it all the time: especially middle management is often more focused on keeping everybody “on board,” using motivational strategies (Red). Lastly, external change agents are drawn toward entrepreneurial behavior (White) and didactic approaches (Green) to gain entry into their client's system as they often lack formal power or long-standing relationships. All these preferences might stand to reason for each of the three types of leadership positions, but imbalances in the color spectrum also pose risks. When a group of

managers score low on Yellow, they may pay insufficient attention to checks and balances with regard to power. Similarly, when staff departments score low on Green and White, this may cause them to lack entrepreneurial spirit and neglect their own know-how. Lastly, when external change agents score low on Blue and Red, they may struggle to reconcile professional distance (Blue) and customer intimacy (Red). None of these risks seem coincidental: they can be regarded as part and parcel of the type of leadership position.

### **Additional differentiation studies**

One might note that the results presented above are not based on a representative sample of respondents. A disproportionate number of respondents are middle-aged or older and highly educated: this corresponds with the typical participants in educational programs on change management. In one of the studies (Test 10 in table 2), we teamed up with an ongoing research project that looked at the values and mentalities of the Dutch population as a whole, a study meant to aid in the segmentation of markets (Knoop, de Caluwé and Mulder, 2009). At the time, the researchers used a representative online panel of 80,000 Dutch people between the ages of 18 and 65, all of whom had at least some vocational education. The main segmentation was in terms of status and values, creating eight different clusters, such as “social climbers”, “new conservatives”, and “post-materialists”. Out of a sample of more than 15,000 people about 4,000 people also filled out the color questionnaire. In this research we looked at the prevalence of clear sympathies or antipathies based on either a particularly high or low color score ( $>0.5$  sd). On average people had about three “extreme” scores in their overall profile, bringing the total to about 12,000 scores. Sympathies and antipathies for all the colors were evenly spread: each accounted for between 9% and 11% of the 12,000 scores. On a population level, this meant 29% of people showed sympathy for Yellow, 27% for Blue, 25% for Red, 26% for Green, and 26% for White, and 31% showed antipathy for Yellow, 30% for Blue, 32% for Red, 26% for Green, and 32% for White. In 51% of the cases people had one dominant color. When another color scored high as well, those combinations showed a pattern: Yellow and Blue scores were paired often (8%), as were Red and Green (6%), and Green and White (6%). Other combinations scored much lower. Correlations in terms of gender, education, age, or leadership position in this study were similar to what the Pietersen (2013) study showed. For instance, men score higher than women on Yellow and Blue and lower on Green, White, and Red.

The study also showed something new: that the color preferences differ markedly between different segments of the population. Without going into the specifics of the segmentation model, a few examples can illustrate this point. The study shows that people who are part of the economic or cultural elite or on their way to becoming part of it (“new conservatives,” “cosmopolitans,” and “social climbers”) preferred colors that advance or maintain a dominant position: they had much higher scores on Yellow and Blue than any

of the other five population segments. In contrast, those who question the status quo or turn away from it based on their critical or postmodern outlook (e.g. “postmodern hedonists” and “post-materialists”) veered towards the other end of the spectrum and had high White scores.

We are often asked about differences in color preferences based on nationality or culture. Most of our experience is based on Dutch audiences using a Dutch test or different types of international audiences using an English test. With regard to the latter, the findings do not appear markedly different (see table 2) but they also do not distinguish between specific nationalities. There are two small samples related to Russia (Test 17 in table 2, given in Russian) and China (Test 18 in table 2, given in English). We observe higher scores on Yellow and lower on White in those two populations in comparison with the Dutch and international scores. We think it is premature to see this as proof of a contrasting profile in these parts of the world. This would require further study.

### **USE AND INTERPRETATION OF THE TEST**

The test is useful as a quick “mirror” of one’s own belief systems about change. In combination with the color model, change agents can use the test to assess to what extent they are making good use of the full spectrum of approaches to change. They can reduce their blind spots, temper their antipathy to certain change preferences and explore new avenues of change that were previously out of bounds. They can also discuss the viability of different (color) viewpoints and approaches to specific issues with others and match people’s capabilities to specific undertakings. These applications become truly powerful when the test triggers people to explore a meta-model of change like the color model more thoroughly. We regard the questionnaire as an entry point to the underlying change theory.

We notice four types of discussions that are often triggered by test scores. A first discussion concerns the extent to which a test score corresponds to change agents’ behavior. The test mirrors what people think, and not necessarily how they act. The two can contrast for several reasons. Some people have a hard time translating their beliefs into action because they lack the capability to do so. In this case, the contrast illuminates possible learning goals for one’s own development. Other people find it hard to act according to their belief system because the issues they work on do not warrant it: the contrast then points to possible career goals, if one wants their work to fit their change preferences. In both of these situations the gap may be uncomfortable but not dysfunctional: people can still act within their competences and in a way that fits the issues. A third explanation is more problematic. When the test mirrors an “espoused theory” rather than the “theory in use” (Argyris, 1998), it may be that people are unaware of how they act, let alone the consequences of their actions. Such a gap can only be

bridged by acknowledging that one's actions show one's true beliefs more accurately than one's speech. Through taking other people's feedback seriously, this veil can be lifted, demonstrating the value of interpreting the scores with a group of colleagues.

A second discussion centers on the benefits and drawbacks of one's color profile. We use the average profile (see table 3) as a reference point for this. People often debate the value of having a balanced profile (close to average) versus a more narrow profile, in which one or two colors are dominant. A broad profile can allow for style flexibility, switching to different approaches where and when needed. This can be useful for those in middle-management positions, where different types of issues arise, none of which can be ignored or easily delegated. In contrast, external consultants can be more selective about the types or clients of issues they engage: specialization allows for a more narrow profile. As each of the colors represents different schools of thought, it is hard to become highly skilled in all five colors in one lifetime. The test scores can lead one to ponder the pros and cons of flexibility versus specialization. We are inclined to speak in favor of specialization where possible: the power of any color's approach is partly determined by the credibility and competence of the change agent. Specialization allows for more "color depth," which benefits change effectiveness.

This last remark is related to a third discussion about collaboration with others who have contrasting profiles. The test can spur conversations about differences and similarities in groups. It can help people face problems with cooperation, facilitate mutual acceptance, and complement each other's qualities. Where a group is able to do so, the test may help them deal effectively with multifaceted issues. A recurring insight is that the existence of contrasting beliefs or values within groups does not determine if they perform well or not. Groups seem to only benefit from diversity when they are able and willing to deal with the tensions it creates (e.g., Shaw & Barret-Power, 1992). There is often a similar debate about whether or not it is beneficial to have a profile that is aligned with an organization's culture. Here, a similar conclusion is often reached: a color profile that contrasts with one's surroundings can allow one to bring something to the table that others don't. It can thus be an added value. However, if one is not able to deal with possible tensions that arise from this difference, such contrasts will fail to bear fruit.

A fourth discussion is about self-fulfilling prophecies. When people have a color preference, they may have an inclination to choose corresponding approaches, gain experience, and build their competence, which in turn reinforces their color preference. Thus a "competency trap" may be created, where they cannot escape that part of the color spectrum even when it is most needed (Levitt & March, 1988). This argument makes sense to the extent that there is sufficient pre-existing competence in that color to allow for some success. In many organizations Blue, Red, and Green approaches are sufficiently widespread to allow for such pre-existing competence. We find it intriguing that the

average score on White is twice as high as the score on Yellow (see table 3), with the other colors hovering in between. This might be explained by White being more fashionable than Yellow in people's imagination: it seems that many like to embrace ideals of self-direction, innovation, and entrepreneurialism (White) more than the commonly disparaged reality of power games and politics (Yellow). This is partly due to "persuasive language," in which the upside of the White world is exaggerated and beautified: every message may be tweaked to form, strengthen, or change the response of others in a desired direction (Aarts & van Woerkom, 2008). The downside of the White world—the hard work that goes into it, its limits in terms of effectiveness—these can easily escape attention especially when there is little pre-existing competence or past experience. This forms an obstacle to pulling off White-print change effectively, thus allowing it to stay more popular in our thoughts than in our actions.

### **CONCLUDING REMARKS**

Our test of points in its final version is a reliable and valid instrument to measure individual change preferences. The average scores and standard deviations are robust and stable over time, and individual scores can easily be compared with the average of demographic segments. The test can give rise to discussions that aid professionalization and collaboration. In our view, the questionnaire is primarily an aid to reflection: the scores are meant to stimulate discussion about their interpretation rather than to be a definitive answer about one's style. We are inclined to regard this limitation as strength rather than a weakness.

The test results reported here give rise to possible new inquiries. One avenue is a more in-depth analysis of the contrast in change preferences between different population segments or cultures. Another avenue is a more in-depth study of what underlies the contrasts between paradigms of change. We were pleasantly surprised by the clear differentiation of the colors that resulted from multidimensional scaling, but our explanation of the two dimensions is by no means the end of the discussion. It could be worthwhile to explore this further and to contrast it with other studies in which change paradigms are analyzed in two dimensions (e.g., Huy, 2001; Higgs & Rowland, 2005).

### **ACKNOWLEDGEMENTS**

Many people contributed to the development of the questionnaire, especially master students, as is clear by the included references. We want to acknowledge B. Pietersen specifically.

## LINKS TO THE TEST



Color test for change agents in English:  
<http://tg.quaestio.com/survey/qst/COLORSCAN>  
(retrieved on January 1st, 2015)



Color test for change agents in Dutch:  
<http://www.twynstragudde.nl/kleurentest>  
(retrieved on January 1st, 2015)

## REFERENCES

- Aarts, N., & Woerkom, C. van (2008). *Strategische communicatie: principes en toepassingen* [Strategic communication: principles and applications]. Assen: Van Gorcum.
- Abbas Zaidi, J. (2013). *Unlearning for Change? Effects of obviously hidden power play, canny politics and deafening silence of managers on Organisational Unlearning. A comparative study in three ex-socialist bloc countries*. Non-published research paper.
- Argyris, C. (1998). *Knowledge for action: A guide to overcome barriers to organizational change*. San Francisco: Jossey-Bass.
- Beer, E., & Nohria, N. (Eds.) (2000). *Breaking the code of change*. Boston, MA: Harvard Business School Press.
- Blinkhorn, S., Johnson, C., & Wood, R. (1988). Spuriouser and spuriouser: the use of ipsative personality tests. *Journal of Occupational Psychology*, 61, 153–162.
- Caluwé, L. de & Vermaak, H. (2003): *Learning to Change. A guide for organization change agents*. Seven Oaks: Sage Publications.
- Caluwé, L. de & Vermaak, H. (2004): Change paradigms: an overview. *Journal of Organization Development*, 22, 4, 9-18.
- Cooper, C.R. & Schindler, P.S. (2008) *Business Research Methods* (10th ed.) Boston, MA: McGraw-Hill
- Higgs, M., & Rowland, D. (2005). All changes great and small: Exploring approaches to change and its leadership. *Journal of Change Management*, 5(2), 121–151.
- Hosking, D. (2006). Not leaders, not followers: A post-modern discourse of leadership processes. In B. Shamir, R. Pillai, M. Bligh, & M. Uhl-Bien (Eds.), *Follower-centered perspectives on leadership: A tribute to the memory of James R. Meindl* (pp. 243–264). Greenwich, CT: Information Age Publishing.

- Huy, Q.N. (2001). Time, temporal capability and planned change. *Academy of management review*, 26(4), 601–623.
- Kessener, B., & Termeer, C.J.A.M. (2006). Organiseren van diepgaand leren: Veranderen als reflexief betekenisgeven [Organizing deep learning: Change as reflective sense making]. *Management & Organisatie*, 3/4, 236–250.
- Knoop, L., Caluwé, L. de & Mulder, M. (2009) Zoeken naar synergie tussen veranderkunde en marketing: een verkennende studie naar overeenkomsten in achterliggende waarden van een veranderkunde en een marketingcommunicatiemodel. [Searching for synergy between change management and marketing: exploring the commonalities in underlying values between two metamodels] *Management en Organisatie*, 6, 20-38.
- Kruskal, J.B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.
- Lankreijer, B. (2007). *Validation: kleurentest* [Validation: color test] Research paper. Amsterdam: Vrije Universiteit.
- Levitt, B., & March, J.G. (1988). Organizational learning. *Annual Review of Sociology*, 14, 319–340.
- Lewis, M.W. (2000). Exploring paradox: Toward a more comprehensive guide. *Academy of management review*, 25(4), 760-776
- Loo, R. (1999). Issues in factor-analyzing ipsative measures: The Learning Style Inventory (LSI-1985) example. *Journal of Business and Psychology*, 14(1), 149–154.
- Martins, Dias, S. (2000). *Een exploratief onderzoek naar het ‘denken over verandering in vijf kleuren’ binnen Twynstra Gudde* [An explorative study of ‘thinking about change in five colors’ at Twynstra Gudde]. Master’s thesis. Rotterdam: Erasmus Universiteit.
- Meyerson, D.E. (2003). *Tempered radicals: How everyday leaders inspire change at work*. Boston, MA: Harvard Business School Press.
- Nistelrooij, A. van, Caluwé, L. de & Schouten, N. (2007) Management consultants’ colorful ways of looking at change. An explorative study under Dutch management consultants. *Journal of Change Management*, 17, 3-4, 243-254.
- Oort, M. (2006). *Preferences for organizational change*. Master’s thesis. Amsterdam: Vrije Universiteit.
- Pietersen, B. (2013) *How valid are the questionnaires on the colors of change by Caluwé and Vermaak? An extensive quantitative analysis of the framework*. Master thesis. Amsterdam: Vrije Universiteit.



- Saville, P., & Wilson, E. (1991). The reliability and validity of normative and ipsative approaches in the measurement of personality. *Journal of Occupational Psychology*, 64(3), 219–238.
- Scherer, A.G., & Dowling, M.J. (1995). Towards a reconciliation of the theory of pluralism in strategic management: Incommensurability and the constructivist approach of the Erlangen school. *Advances in Strategic Management*, 12, 195–247.
- Schön, D.A. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Sekaran, U. & R. Bougie, R. (2009). *Research methods for business. A skill building approach*. (5th ed.). West Sussex (UK): Wiley & Son
- Shaw, J.B., & Barret-Power, E. (1998). The effects of diversity on small work group processes and performance. *Human Relations*, 51(10), 1307–1325.
- Smith, W., & Lewis, M. (2011). Toward a theory of paradox: A dynamic equilibrium model of organizing. *Academy of Management Review*, 36(2), 381–403.
- Tummers, M. (2009). *Validation of the colors of change: A meta-analysis of 1.737 of change agents*. Master's thesis. Amsterdam: Vrije Universiteit.
- Velde, E.G., van der, Jansen, P.G.W. & Anderson, N.A. (2008). *Guide to Management Research Methods*. (2nd ed.) Oxford; Wiley Business Economics
- Ven, A.H. van de, & Poole, M.S. (1995). Explaining development and change in organizations. *Academy of Management Review* 3, 510–540.
- Vermaak, H & Caluwé, L. de (2015). *Creating a colorful model of change. A case study of theory development*. Paper Academy of Management Annual Conference, Vancouver
- Wierdsma, A. (2007). A methodology for increasing collective competence: A context for co-creative change. In J. Boonstra & L.de Caluwé (Eds.), *Intervening and changing: Looking for meaning in interactions* (pp. 243–260). Chicester: Wiley.
- Xu, Y. (2011). *A research of investigating cultural differences between Chinese and Dutch in the working environment*. Master's thesis. Amsterdam: Vrije Universiteit.

TABLE 1. The five colors at glance

	<b>Yellow-print</b>	<b>Blue-print</b>	<b>Red-print</b>	<b>Green-print</b>	<b>White-print</b>
Something changes when you...	bring common interests together	think first, then act according to a plan	stimulate people in the right way	create settings for collective learning	create space for spontaneous evolution
in a/an....	power game	rational process	trading exercise	learning process	evolutionary process
and create....	a feasible solution, a win-win situation	the best solution, a brave new world	a motivating solution, the best 'fit'	a solution that people develop themselves	a solution that releases energy
Interventions such as...	forming coalitions, changing top structures	project management, strategic analysis	assessment & reward, social gatherings	gaming and coaching, open systems planning	open space meetings, self-steering teams
by...	facilitator who use their own power base	experts in the field, project managers	HRM experts, a managers who coaches	facilitators who support people	persons who use their being as instrument
aimed at...	positions and context	knowledge and results	Procedures and inspiration	setting and communication	patterns and meanings
The result is..	unknown and shifting	defined and guaranteed	outlined but not guaranteed	envisaged but not guaranteed	unpredictable but not aimless
safeguarded by...	decision documents and power balances	benchmarking and ISO systems	HRM systems and healthy relationships	a learning organization	selfmanagement and dialogical quality
The pitfalls lie in...	dreaming and lose-lose	ignoring external and irrational aspects	smothering and conflict avoidance	excluding no-one and lack of action	superficial understanding and laissez faire

TABLE 2. Different tests and samples studied

Test	Type and version	Type of test	Authors/ year	Language	Paper or electronic	Type of sample	Num-ber of respondents (N)	Main research finding				
								Yellow	Blue	Red	Green	White
1.	0.1	Scale test (agree/neutral/dis-agree) (30 items)	Martins Dias (2000)	NL	P	Change agents/ students	50	Average				
								Yellow 2.90	Blue 1.66	Red 3.30	Green 3.88	White 4.36
								Proportional average*)				
								Yellow 17	Blue 10	Red 20	Green 23	White 26
2.	1.1	Forced choice between A and B; (30 items on how people think and 30 items on how people act)	De Caluwé & Vermaak (1999)	NL	P	Readers	-	Data lost				
3.	1.2	Forced choice between A and B; (30 items on how people think and 30 items on how people act)	De Caluwé & Vermaak (2003)	E	P	Readers	-	Data lost				
4	1.3	Forced choice between A and B; (30 items on how people think and 30 items on how people act)	Website Twynstra Gudde (2005)	NL	E	Visitors	-	Data lost				
5.	2.1	Test of points (10 items)	Martins Dias (2000)	NL	P	Change agents/ students	50	Average				
								Yellow 13	Blue 11	Red 14	Green 18	White 24
								Proportional average*				
								Yellow 16	Blue 13	Red 17	Green 22	White 29
6.	2.2	Test of points (10 items)	Oort (2006)	NL	p	Change agents/ managers/ support staff	2.688	Average				
								Yellow 12	Blue 13	Red 18	Green 17	White 20
								Proportional Average*				
								Yellow 14	Blue 16	Red 22	Green 20	White 24
								Indications to improve some items				
7.	2.3	Test of points (12 items)	Lankreijer (2007)	NL	P	Change agents/ managers	280	Indications to improve items				
8.	2.4	Test of points (12 items)	Website Twynstra Gudde (2007)	NL	E	Visitors	36.664	Average				
								Yellow 15	Blue 20	Red 20	Green 19	White 22
9.	2.4	Test of points (12 items)	Tummers (2009)	NL	P	Change agents/ managers/ support staff	1.737	Average				
								Yellow 13	Blue 18	Red 18	Green 20	White 25
10.	2.4	Test of points (12 items)	Knoop et al. (2009)	NL	E	Representative sample of the Dutch population	4.086	Dominant colors are equally distributed among the Dutch population				

11.	2.4	Test of points (12 items)	Website Twynstra Gudde (2009)	E	E	Visitors	18	Average				
								Yellow 14	Blue 20	Red 20	Green 23	White 20
12.	2.5	Test of points (12 items)	Website Twynstra Gudde (2010)	NL	E	Visitors	47.237	Average				
								Yellow 13	Blue 20	Red 20	Green 20	White 23
13.	2.5	Test of points (12 items)	Website Twynstra Gudde (2010)	E	E	Visitors	883	Average				
								Yellow 15	Blue 20	Red 21	Green 21	White 20
14.	2.4	Test of points	Pietersen (2013)	NL	P	Managers/ change agents/ support staff	2.702	Average				
								Yellow 13	Blue 19	Red 18	Green 20	White 26
15.	2.5	Test of points	Pietersen (2013)	NL	P	Managers/ change agents/ support staff	665	Average				
								Yellow 13	Blue 19	Red 18	Green 20	White 26
16.	2.5	Test of points	Pietersen (2013)	E	P	Managers/ change agents/ support staff	91	Average				
								Yellow 15	Blue 17	Red 20	Green 20	White 24
17.	2.5	Test of points	Abbas Zaidi (2013)	Rus- sian	P	Managers and workers in Russia	243	Average				
								Yellow 22	Blue 20	Red 20	Green 19	White 15
18.	2.5	Test of points	Xu (2011)	E	P	Chinese people working in NL	50	Average				
								Yellow 21	Blue 21	Red 20	Green 16	White 19

\* Proportional average is the adjustment of test results of various tests to a total score of 96 as is the case in the 12 items test of points

TABLE 3. Average scores in three samples with the latest versions of the test

	Test of points, version 5 (Test 15 in table 2) (Pietersen, 2013) (N=665)		Test of points, version 4 (Test 14 in table 2) (Pietersen, 2013) (N=2702)		Test of points, version 4 (Test 9 in table 2) (Tummers, 2009) (N=1737)	
	Average score	SD	Average score	SD	Average score	SD
Yellow	13	6.56	13	6.55	13	6.6
Blue	19	9.35	19	9.44	18	9.2
Red	18	6.69	18	6.60	19	6.3
Green	20	8.21	20	7.69	20	7.4
White	26	8.50	26	8.69	25	8.2

FIGURE 1. Results of the multidimensional scaling analysis (MDS) showing clear separation of color preferences in the test results.

