Documentation for description and evaluation of the Diversity Icebreaker

Bjørn Z. Ekelund, Piotr Pluta, Harald K. Ekelund

Makes people reflect, ask, listen and talk

The Diversity Icebreaker® is certified by the DNV Business Assurance as a team and development tool for use in workshops

Table of Contents

Foreword

This document is to accompany the documentation sent by Human Factors AS and help the evaluators to orient themselves in regard to the materials and sources provided.

The document corresponds to particular sections of the DNV-evaluation form. It is intended to be a systematic presentation of different aspects of the Diversity Icebreaker as a product and a psychological concept.

A considerable part of this document is dedicated to the studies supporting the validity of the Diversity Icebreaker. Best effort is made to present these studies in a systematic way, but given the vast amount and diversity of the studies, there is some variety in the presentation that could not have been avoided. The diversity of authors and approaches in these studies is a result of an important part of our quality and business philosophy, which has been to openly share our experiences (via our homepage, publications and presence on over 20 conferences) and, through this transparency, also invite others to comment and discuss our work.

Some studies will be presented to a greater extent and detail as they have not been published elsewhere; other will only be briefly described with reference to the complete publication and/or the provided attachments. This document also reports on-going research, indicates areas for improvements, and lists references.

In the text, there are cross-reference links to different sections of the document to facilitate navigation.

What is the Diversity Icebreaker (Section 1: General information)

The Diversity Icebreaker (DI) is a training and development concept for individuals and teams based on a questionnaire measuring preferences in communication, teamwork and problem-solving styles.

Using three categories termed Red, Blue and Green, the questionnaire does not assign individuals to single types, but provides a profile of preferences across these three categories. Red stands for a social orientation, with feelings and personal communication in the centre; Blue is a preference for analytical thinking, figures and details; and Green is related to broad ideas, big picture thinking and making connections (for most complete descriptions of the categories refer to the *Diversity Icebreaker Personal Workbook*, pp. 7-9 and to the *Red, Blue and Green preferences for communication and interaction* section below).

The DI questionnaire and the categories of Red, Blue and Green has been developed primarily by Bjørn Z. Ekelund of Human Factors in Norway during the years from 1995 and onwards.

The classic DI workshop begins by asking individuals to map their own preferences using this questionnaire. Then in dynamic group work, participants are encouraged to discuss their own preferences and their perceptions of others, creating an environment to reflect and to identify ways to cooperate more effectively.

The Diversity Icebreaker is typically used in six areas: a) team, leadership, project work and innovation; b) cross-cultural trainings; c) kick-offs at large events; d) diversity management, e) communication trainings and conflict management; f) self-understanding.

The classic Diversity Icebreaker workshop

By the "classic" DI workshop is what we understand the scenario, fixed structure and guidelines for conducting a workshop, based on the DI questionnaire; which are described in the *Diversity Icebreaker User Manual* and demonstrated in the instructional DVD (attached).

The workshop scenario consists of four subsequent stages. In the first one, the participants fill in the questionnaire and score the results. They obtain results on three dimensions: Red, Blue, and Green – labels that bear no meaning for them as the preferences and colours had not been previously explained. In the second stage, the participants are dealt in three groups of same size (ones who scored highest on Red, Blue and Green) and asked to work together to answer two questions:

> "What are the good qualities of your own colour in interaction with others?"

and:

> "What are the qualities of the two other colour groups in interactions they have with each of the other groups?"

In the third stage, the groups are asked to present the results. The way how the participants in one group perceive their own colour is contrasted with how the other two groups perceive it, and attention is given to the processes of social construction taking place when the meaning of Red, Blue and Green is being negotiated.

The fourth stage is a learning process which is initiated by asking the participants a question:

"What have you learned from the time you started filling in the questionnaire, until now?"

The person leading the workshop (consultant or facilitator) then moderates the discussion in order to reach a number of learning points and illustrate them in the workshop's retrospection. Examples of these learning points are: people are different; it is easy to work together with people that are similar, but we need differences in order to reach our objectives and be creative; we all have different qualities (colours) and we can draw on our less predominant preference to connect better with others, etc. (refer to the *Diversity Icebreaker User Manual* and DVD for more examples). These learning points seem to be relatively consistent across groups and consultants, and are often arrived at spontaneously by the participants.

Often, a fifth stage is added to the workshop, which is concerned with developing specific practices for the future. This emerges as a salient follow-up of the learning points from the fourth stage.

History of Red, Blue and Green

1994: The creation of Red, Blue and Green

In December 1994, Human Factors AS started a five-year consultation project with a client within the energy industry. The area of consultation was "how to make people reduce energy consumption". The categories of Red, Blue and Green as communication strategies emerged early in this project and became the platform for specific market communication and consultation towards different consumer segments. A full description of the challenges facing the client, the design process, implementation and evaluation of this engagement have been presented in Ekelund's dissertation for the MBA in 1997 at Henley, London (Ekelund, 1997). Here, I will focus more on the process, in which the categories of Red, Blue and Green emerged, with a discussion of the consequences of this method in relation to psychological research, as well as on its practical use.

The client was involved in marketing and consulting concerning the reduction of energy consumption in private households. Together with a group of marketing and public relation companies, Human Factors AS was asked to redesign their communication strategies so that they could reach out for new target groups. In order to define these new groups and suggest ideas for better communicating with customers representing these market segments, customers were invited, in December 1994, to take part in focus groups working on this general communication challenge -"How to communicate in order to make people reduce energy consumption?" Twenty-seven customers in four groups took part in group work. The work was organised according to the principles of brain-writing presented in the work of VanGundy (1981), combined with ideas from the qualitative method for conceptualizing unstructured material as described in Strauss & Corbin (1990). The customers produced 161 ideas in total and were then asked to group the ideas based on "what belonged to each other - and what was different". Three main groups of ideas emerged in all the four focus groups, and they comprised 121 ideas out of the overall 161. The 40 other ideas that did not fit into the three main groups varied in structure and content to an extent that made it difficult to categorize them in a meaningful way, and for this reason it was impractical to prioritize them as targeted goals in the market communicative initiative. Of the three main categories, one was defined as being economically motivated, one as environmentally motivated, and the last one as motivated by social factors. The groups were assigned the three-color nicknames because of the similarities with the political colours in the Norwegian political party structure: Blue (the conservative side, more concerned with economic

conditions), Red (the social democratic / socialist side, more concerned with social welfare) and Green (the environmentalists, more willing to take a global perspective).

The marketing and communication campaign was designed to follow the sequence described below:

a) Attract attention through different market campaigns in newspapers, where Red vs. Blue vs. Green communication strategies varied in content and colour background. It was not intended that the colours themselves should convey any meaning. They only identified the strategies for internal and external communication. The market communication ended by inviting people to make contact with the client in order to get advice and tools for reduction of energy consumption.

b) Advisors from the client organization were trained in communication in order to reinforce the customers' seeking-contact-for-more-information behaviour in order to build their self-confidence as information seekers and competent problem solvers in this area.

c) The communication training focused also on how to give advice for technological and behavioural change in the Red vs. Blue vs. Green strategies due to what seemed to trigger the customer's interest and build his/her confidence in future behavioural change.

d) Written material was given to clients where they could read guidelines and information themselves and self-select arguments from a Red vs. Blue vs. Green perspective.

In retrospect, the whole process of attracting interest by differentiated Red vs. Blue vs. Green adverts, establishing a relationship, reinforcing the relationship and then introducing differentiated Red vs. Blue vs. Green information within the established communicative relationship, seems to be a complex behavioural-cognitive–attitudinal process, aligned with communicative strategies in order to influence the behaviour of others.

A central aspect that has not been highlighted in the original work in 1997 (Ekelund, 1997) is that the categories of Red, Blue and Green emerged from a process where randomly invited customers (a random sample of individuals who were not experts in marketing, social communication, or energy consumption), were involved. The three categories emerged in what Moscovici termed the process of common sense categorization, in opposition to reified scientific methods (1984), like factor analysis where each factor is intended to capture the entire variance of a given variable or facet. On the one hand, this may lead to the categories being easier to understand and various behaviours easily ascribed to one of the colours by the seminar's participants. On the other hand, it might be problematic to confirm the three colours as orthogonal, separate factors in the analysis of variance. Since only 121 out of 161 ideas were grouped belonging to the Red, Blue or Green categories, the remaining 40 were probably examples of behaviour that does not fit into the structure. The total picture of variance is thereof not captured. A consequence of this common sense categorisation is a conceptual structure that might be difficult to comply with classical statistical analysis strategies usually applied to results of personality questionnaires, such as factor analysis. (We direct reader to the *Factors structure* section below for more information regarding this issue.)

1997: The construction of the questionnaire

In in 1997 we designed a questionnaire that would make it possible to identify levels of Red, Blue or Green among the different customers interacting with the company at different times and places. Two psychological traditions that shared similar ideas of "how to communicate in order to change the behaviour of the other" were recognised in the theoretical work of creating adverts and

communication strategies. One was the tradition of learning and teaching styles (Honey & Mumford, 1992, Gardner, 1993), and the other one was based on team role concepts where Belbin´s and Margerison & McCann´s work was perhaps best known and often used in Europe (Belbin, 1981; Margerison & McCann, 1991) while MBTI was widely applied in the USA (Matthews & Deary (1998). Margerison & McCann's model gives practical advice on communication styles based on different typologies (McCann, 1988). These perspectives influenced both the campaign strategies and the creation of the first version of the questionnaire used as a market segment identifier. In the development of the first Diversity Icebreaker questionnaire one hundred questions were picked from already established concepts of personality, interactional preferences and team roles. The questions were all formed into a Likert scale format. One hundred persons answered the questionnaire. Fifteen items on each colour were identified through cluster analysis performed in a simple form where items were chosen in and taken out by hand. The criteria for what was selected in or out were internal reliability and face validity of the items. The internal reliability measured with Cronbach's alpha of the dimensions of Red, Blue and Green in this first study with N=100 and using Likert scales were medium; Green α=.62, Red α=.66 and Blue to α=.70 (Hegge, 1997).

In the first questionnaire it was decided not to use the Likert scale but rather ask respondents to prioritise one out of three items, thus following a classical ipsative format. The three items were taken from the group of items belonging to Red, Blue or Green according to the analysis described above. Which three items to present in opposition to each other, was decided by picking out items that had the same type of content or verbal structure (for example, Red: I show my feelings; Blue: I am practical-minded; Green: I often try new things). The scoring results varied between 0 and 15 on each of the three dimensions. The sum for each respondent was 15 due to the ipsative format.

1998: Launched as an alternative team-role concept.

The questionnaire was included in the training manuals (Ekelund & Jørstad, 2002) for the concept of Team Climate Inventory (Ekelund & Jørstad, 1998) as a simpler team-role concept and an alternative to TMS (Margerison & McCann, 1991) and Belbin (1981). This strengthened the use of the concept in team and organisational development where preferences for interaction and distribution of tasks were the main focus. The chapter on managing diversity in cross-professional teams is aligned with this perspective (Ekelund, 2009).

2004: A separate questionnaire with training material

Following these publications and the consultants' use of the questionnaire we received positive customer feedback and requests for more advice concerning the use of both the questionnaire and its application in seminars. In response, we published a brochure in 2004 integrating the questionnaire with training guidelines. This brochure also featured a structured description of a process whereby the participants are asked, in groups of the same colour, to put the ideas they share about their own colour, as well as ideas about the other two colours, on flip-chart in order to share them later with other groups. In this way the gap between actor and observer perspectives, and personal vs. social identity (Tajfel & Turner, 1979) could be pedagogically utilized (Jones & Nisbett, 1972). These steps were later to become the first three stages of what has been named "the classical Diversity Icebreaker" seminar, described in different papers and books in 2006 (Ekelund & Langvik, 2006, Ekelund & Langvik, 2008, Ekelund, Davcheva & Iversen, 2009, Ekelund, 2010).

2004: First revision of the questionnaire into a partial ipsative format

The ipsative scoring format described above does not afford the possibility of ranking between the two items that were not preferred. In 2004 we introduced a version of the ipsative format, in which the respondents were asked to spread the total of six points between three different items opposing each other. This gave each item a potential of getting between 0 to 6 scores, which increased the possibility to run more advanced statistical analysis, e.g. different variance analysis. This particular scoring format was named "partially ipsative".

This format is ipsative in the sense that the scores in the Red, Blue and Green dimensions are dependent on each other. If a person scores very high in Blue, the scores in the other two dimensions will be correspondently lower. This forced choice format is especially good for intrapersonal comparison, i.e., the individual's score in one dimension is compared against his or her score in the other dimensions (Langvik, 2006) The psychometric characteristic of ipsative scales differs substantially from that of normative, i.e. traditional scales since ipsative scales force dependency among responses given by the individual (Nysæter et al, 2009) . Ipsative scores in factor analysis are claimed to produce artificial bipolar factors and for that reason factor analyses of ipsative data are often dissuaded (Dunlap & Cornwell, 1994). This format, however, opened up opportunities for variance analysis used for documenting and improving reliability, as well as made it possible for us to do empirical validation studies in the years to come. More empirical comparisons of the use of the partially ipsative format compared to the Likert scale format are needed to explore and ensure the quality of use of different statistical procedures.

2005: Second revision of the questionnaire

The version used from May 2004 until June 2005 consisted of 15 items for each dimension of Red, Blue and Green. A study of internal consistency in 376 respondents showed that one of the questions was negatively correlated with the dimensions (Ekelund & Langvik, 2008). From July 2005 the questionnaire was reduced to 14 items per dimension and this is the version in use today (2012). The internal consistencies measured with Cronbach's alpha were for Red $\alpha=.81$, Blue $\alpha=.82$, and Green $\alpha=.75$ (Ekelund & Langvik, 2008).

2005: First data analysis with empirical validation

Since 1995 the concept has been described in relation to different theoretical constructs, but the first empirical validation studies began in 2004 following the introduction of the six-point partially ipsative format. The validation process was mainly focused on reliability and construct validity with convergent and divergent validity of the Red, Blue and Green dimensions compared to other psychological assessments. The construct validity processes described by Cronbach & Meehl (1955) and Campbell & Fiske (1959) guided us in creating meaning to the categories. Face and content validity of the dimensions were partly neglected due to the lack of theoretical precision in the creation of categories and the improvised way of creating the specific items in the questionnaire in 1997 (Hegge, 1997). The consequential validity (Messick, 1995) of the categories had been documented in the area of marketing (Ekelund, 1997).

Measurement & Scoring (Section 3)

The standard Diversity Icebreaker questionnaire
The Diversity Icebreaker in its primary form is a one-side, self-scoring, paper questionnaire consisting of fourteen questions.

Each question consists of three statements, e.g. Question 2: *Jeg liker å tenke logisk – Jeg kan lett bli så oppslukt av en idé at praktiske detaljer oversees – Jeg trives med å omgås folk som jeg ikke kjenner godt*.

The questions are fitted with partial-ipsative format, where the respondent has to distribute six points (ticks) between the three statements per question, so that the more ticks one puts on an statement, the better this particular statement describes him or her. All combinations that sum up to six are accessible, e.g. 2+2+2, 2+3+1, 3+3+0, etc. Refer to the *First revision of the questionnaire to the partial-ipsative format* for more information about the scoring format.

The questionnaire is made of two-sheets of carbon copy paper – after filling out the questionnaire, the respondent tears off the first page. On the second page, questions and statements appear in the same order and location on the sheet, the difference being that they are coloured in red, blue and green. The carbon-copy paper makes the ticks written with a bullet-point pen (preferably) visible on this page as well. The respondent can then calculate his or her result by adding all the ticks per colour, i.e. summing all the ticks he or she had placed on the red fields, blue fields and the green fields.

After scoring the questionnaire the respondents obtain results on Red, Blue and Green – which the processes in the DI workshop are based on afterwards. The Red, Blue and Green scores should add up to 84 (which is communicated to the participants as a mean of "quality control").

On the bottom of the first page of the questionnaire, there is space to enter demographic data (sex, age, organization) for research or norm-building purposes where applicable.

It takes between 10-15 minutes to fill out and score the questionnaire.

The current version of the Norwegian paper questionnaire is V11 (December 2012).

Likert scale versions
There are Likert scale adaptations of the DI questionnaire used both in paper as in an on-line version (EasyFact platform employed) for research purposes. It is noted where this or the standard, partial-ipsative scoring format was used when different validation and reliability studies are presented in this document.

Supply, Conditions of use and Costs (Section 5)

Conditions of use

We require no certification or a license for using the Diversity Icebreaker questionnaire in developmental processes with groups or individuals – consultants and facilitators purchase the number of the questionnaires they require.

However, in order to send the materials, we require that an individual purchasing and thereafter using the tool registers him- or herself in our database and makes a personal declaration. At the registration, a person is asked to do the following: confirm that he or she has the adequate competences to conduct group processes; read the User Manual thoroughly and get acquainted with the available material before leading the session; do not let another, non-registered person use the questionnaire; and provide contact information. During the registration process one is also asked to describe his or her experience with the tool. We evaluate each registration before we expedite the order, and we contact the buyer if we are unsecure about their aptitude to use the tool.

A person is asked to register him- or herself for the first time when placing the order for the questionnaires, on the ordering-form website: www.divorder.no . The registration form is available on www.dibruker.no website.

NOTE: Both in the material (the Profile Folder, back page) and on the Diversity Icebreaker website it is noted that the Diversity Icebreaker is not a tool for selection of employees and shall not be used for such.

The Diversity Icebreaker resources

USER MANUAL

After ordering any number of copies of the Diversity Icebreaker and once the registration process has been accepted, the buyer receives an instant access to the User Manual: a practical guide to the classic DI workshop. It describes the workshop step-by-step, introduces alternative ways of managing the process and gives tips on using the tool in different types of groups. The Manual is available in digital version and in print. The printed version is always sent to the first-time users together with the test materials. A consultant or facilitator with experience in leading group processes should feel confident to conduct the workshop after reading this manual. See the *Secure area* below for an overview of these resources/files, exclusively available for DI users.

INSTRUCTIONAL DVD

Additionally, there is an instructional DVD of 32 minutes available, illustrating a real-life DI session led by Bjørn Z. Ekelund. The DVD is available at a low cost. We recommend the DVD for the new users of the tool, as good mean of getting acquainted with the process (especially for those customers without extensive experience in leading group processes). In our experience, most of our customers have seen this DVD. Read more about the DVD here.

BOOK

There is also a book *Diversity Icebreaker. How to manage diversity?* (Ekelund & Langvik, 2008) available for purchase, featuring a broad presentation of the tool together with validation studies and examples of application. Read more about the book here.

PERSONAL WORKBOOK

Either the Personal Workbook or the Profile Folder is provided as a free addition to the questionnaire (a copy per every single copy of the questionnaire ordered; both available for an additional fee of NOK 20, - per copy). The Personal Workbook is available in English and Norwegian in print and in 5 other languages; total of 8 languages downloadable from our website (NOTE: not in the secure, consultant-only area but in the general Material section of our website, which makes it also available for the workshop participants after the session).

The Personal Workbook is a thirty-page booklet containing supplementary information, discussion topics and exercises. Participants can use it individually or with a facilitator in groups. Exercises and questions contained in the Workbook help to summarize and reflect on the learning points, and set them in the context of teamwork, management of diversity, working across cultures etc.

PROFILE FOLDER

The Profile Folder is an A4, three-fold brochure included in the price of the questionnaire. It contains short exercises for individual work as well as concise information about the Diversity Icebreaker. Participants can use it to plot their personal profile based on results from the questionnaire and compare it with normed scores. They can also read about the different preferences and get practical tips on how to improve interaction with others.

BADGES

An aid applicable in the DI workshops and follow up exercises; a colourful pin-up badge (either in Red, Blue or Green). The badges are available in two versions: with or without the text (allowing for personalization). On the text-version an inscription is made, e.g. on a Blue badge "… but also Red and Green" is written (Figure 1 below), illustrating one of the learning points arrived at in the workshop (see the *classic Diversity Icebreaker workshop* section above). One of the purposes of this text is also to emphasize that one is not to be labelled with one and only one preference (which can result in a negative experience of stigmatization).

Figure 1. *The Diversity Icebreaker badges – Green*

Furthermore, we provide the users of the Diversity Icebreaker with a number of easy accessible online resources in order to aid them in conducting the Diversity Icebreaker workshops and other, follow-up processes. This, we believe, is also a mean of supporting the quality-use of the tool.

SECURE AREA

Each person purchasing the questionnaire receives an overview and access to a secured area of our website via a secured link, where the following resources can be downloaded:

- The Diversity Icebreaker User Manual in a digital form (described above)
- PowerPoint presentation – the classic DI workshop (used as an aid in the classic Diversity Icebreaker workshop)
- Excel matrix (for entering data and presenting group profiles on Red, Blue and Green; see the *Local norms* section below)
- PowerPoint presentation – supplementary information (a database of slides reflecting different areas of application of the Diversity Icebreaker)
- PowerPoint presentation – sales presentation (a PPT aimed at aiding to sell the tool's application)
- Sales brochure format A5
- Sales brochure format A4

On this page we also remind that the Personal Workbook and case studies are available for download from our website (in the Material section).

SOCIAL MEDIA

We run different social media sites and we encourage the DI consultants and workshop participants to join them to get updated regarding new materials available, changes and updates, publications and research studies, and events and conferences featuring the Diversity Icebreaker.

These social media sites are: Facebook (featuring short updates and "light content"), LinkedIn group (a place for discussions between the DI consultants), YouTube channel (with testimonials, interviews and presentations about DI) and a blog authored by Bjørn Z. Ekelund (longer posts, with both academic and professional focus).

The Diversity Icebreaker courses and events

We organise a series of different train-the-trainer courses and demonstrations of our tool. Follow this link for the current schedule.

- breakfast seminars of 1,5 h on various topics related to the Diversity Icebreaker (e.g. introduction to the tool, cross-cultural application or follow-up exercises); meant for those already acquainted or proficient in the concept
- demo-workshops; a life session meant as introduction for newbies
- half-day, train-the-trainer seminars (detailed introduction to the classic DI workshop, after which a person is capable to conduct the session knowing different alternatives, etc.)
- two-day course in advance use of the Diversity Icebreaker (an advance train-the-trainer, interactive course focusing on alternatives and challenges when conducting the classic DI workshop, different areas of DI's application, theoretical knowledge and research data behind the tool, discussion of case studies and follow-up exercises)

- *Diversity Icebreaker Forum* (an event where a half-day course is combined with an opportunity to exchange experiences with other DI-users and Human Factors' consultants; it will be held on 18.10.2013)

Prices

As aforementioned, we require no licence fee or certification; all fees are based on "per-copy of questionnaire" basis. One copy of questionnaire can only be used with one individual, one time only. Below is the listing of prices of the questionnaire and each of the other products we sell under the Diversity Icebreaker brand:

1. Diversity Icebreaker questionnaire + Profile Folder or Personal Workbook (in print)
    a. NOK 160, - per questionnaire + VAT
    b. 50% discount for application with students and pupils in the educational sector
    c. 20% large-volume discount (over 100 copies)
    d. Framework agreements for discounts for individual, large customer
    e. + NOK 20, - if one desires both the Profile Folder and the Personal Workbook
    f. The questionnaire is provided for free to external researchers (e.g. universities), in line with the "open innovation model" of our company.
2. DVD
    a. NOK 298, - + VAT.
    b. The film is not available online.

3. Book: *Diversity Icebreaker. How to manage diversity?* by Bjørn. Z. Ekelund and Eva Langvik
    a. NOK 198, - when ordered together with questionnaires and NOK 298,- when ordered alone.
4. Pin-up badges
    a. NOK 10, - + VAT/ piece

Norms, Reliability and Validity of the Diversity Icebreaker (Section 7)

This section will discuss how was the data for the norms for the Diversity Icebreaker assembled and analysed; it will report the reliability studies for the DI scales as well – in its most extensive part – discuss the body of evidence supporting the tools validity.

Norms

The norms for Red, Blue and Green in the Diversity Icebreaker for the Norwegian population as well as international norms are presented in this section. Reader must note however, that the DI norms are not typical norms per se in the context test is being used, i.e. they are not used as cut-off scores for any forms of selection, nor are they used to assign individuals to certain types. They are used only as a reference point and to make comparisons for different levels of Red, Blue and Green between organisations, roles, professions, sexes and cultures. Furthermore, for an individual the norms also give some indication of what "high" and "low" means; and where standard deviation gives one an indication of "how far from the mean" intervals for different scales are (there are different SDs for Red, Blue and Green scales).

The scales in the norms are made by only analysing each dimension at a time – independent of what scores the respondents have on the two other scores. At individual level the single dimension score is high or low compared to the scores on the two other dimensions due to the ipsative format. For this reason the individual cannot use the norms to compare him or herself to "a certain type of person" due to the levels of all three dimensions, like for example *High Blue*, *High Red* and *Very Low Green*.

Norm no 1, N=1378 (Norway, 2006)

We started to gather data systematically for the purpose of building norms in august 2004 when the partial-ipsative format was introduced. We gathered data from all types of workshops where the Diversity Icebreaker was used. Most often they were internal seminars or training programs where participants came from different organizations (project training, leadership training for engineers, psychologists, etc.), but also a few conferences.

Among the data used to establish the Norm no 1, N=1371 (Norway, 2006) about 60% of the data was gathered by the consultants working for Human Factors AS at that time, and 40% were gathered by consultants in organizations that were trained for teamwork and interpersonal training (mainly internal HR-people in large organizations).

Data for this norm were being gathered from 22nd of August 2004 to 23rd of June 2006. These include samples from approximate 33 separate organizations/contexts, estimated to originate from about 80 different seminars or organizational settings. All samples except one (Technical College in Denmark with 13 employees) have been gathered in Norway, with Norwegian organizations. Of all, 78 participants, including the 18 from Technical College in Denmark, about 6% in our sample have a non-Norwegian cultural background. In the Norwegian society this figure is 13% - meaning that it is less non-Norwegians in our sample compared to our society. The sample is a mixture of data from public and private institutions, with higher-degree students totalling to N=100 in the sample.

A sample used is representative, as long as it represents all relevant sub-groups in a given population. Our sample can be considered representative for the intended population, because it is similar to the population in the sense that no systematic exclusion was made in the sampling procedure: In this model the chance of being included in the sample is equal for each sub-group in the population defined as "customers of a I/O and HR-oriented consultation company", i.e. organizations and groups that we normally meet in our consultation work.

There has been no samples that has been taken out that we could have access to in this period, as long as we were sure that the data gathered was scores and registered in a reliable way. We have no reason to assume that our company has a unique skewness in the market that reduces generalizability of the norms to be used in similar contexts. This is something we will challenge later when we see the opportunity to gather market specific norms in our further use of the concept in other countries through partners or subsidiaries.

Mean, standard deviation and gender differences
Means, standard deviations and differences in gender for Norm no 1 are presented in Table 1 below. Analyses of gender differences were significant: Women score higher on Red, t= 12.70, (1030), p<.001, whereas men score higher on Blue, t=7, 83 (1030), p<.001 and Green, t=3.67 (1030), p<.001. There were no significant differences between different age groups for Red, Blue and Green.

Documentation for description and evaluation of the Diversity Icebreaker

Table 1. *Means and standard deviations*

DI factors	All (N=907)	Men (N=354)	Women (N=553)
Blue	M=30.36, SD=7.9	M=32.40, SD=7.9	M=28.82, SD=7.3
Red	M=29.13, SD=7.3	M=25.89, SD=7.3	M=31.42, SD=6.8
Green	M=24.51, SD=7.2	M=25.44, SD=7.1	M=23.80, SD=7.2

Transformation of raw scores – standardised norms

To simplify comparison of individual scores on Red, Blue and Green, standardised norms have been developed. The scores were transformed into 10 categories, based on percentiles: Each of the categories contains 10% of all observations. This transformation presents the person's score in a unit that is more informative and interpretable than the raw score. Table 2 presents the percentile norms, based on the total sample of N=1378.

Table 2. *Norm based percentile scores*

N=1378	Blue	Red	Green
90-100%	>42	>37	>35
80-90%	39-41	34-36	31-34
70-80%	36-38	32-33	29-30
60-70%	34-35	30-31	27-28
50-60%	32-33	29	25-26
40-50%	30-31	27-28	23-24
30-40%	28-29	25-26	22
20-30%	25-27	23-24	20-21
10-20%	22-24	20-22	17-19
0-10%	<21	<19	<16

Table 3 and Table 4 present the percentiles based on the norms for men and women respectively, which allows for gender based comparison.

Table 3. *Percentile scores based on norms for men*

N=621	Blue	Red	Green
90-100%	>44	>34	>35
80-90%	40-43	31-33	31-34
70-80%	37-39	29-30	29-30
60-70%	35-36	27-28	28
50-60%	33-34	26	26-27
40-50%	31-32	24-25	25
30-40%	29-30	23	23-24
20-30%	27-28	22	21-22
10-20%	24-26	18-21	18-20
0-10%	<23	<17	<17

Table 4. *Percentile scores based on norms for women*

N=757	Blue	Red	Green
90-100%	>39	>39	>34
80-90%	36-38	36-38	30-33
70-80%	34-35	34-35	28-29
60-70%	32-33	33	26-27
50-60%	30-31	32	24-25
40-50%	29-29	30-31	22-23
30-40%	26-27	28-29	21
20-30%	23-25	26-27	19-20
10-20%	20-22	23-25	16-18
0-10%	<19	<22	<15

The dominant tendency in gender difference is that women score higher on Red than men do. This is in coherence with the traditional opinion of women regarding how they communicate differently –

being more relationship- and emotional-oriented – and has been confirmed to certain extent in validation studies of the Diversity Icebreaker (see *Convergent internal validity* section).

APPLICATION OF NORMS

When applied for individual self-understanding – it is relevant to decide whether one should use gender specific norms or not. When applied for organizational or team development and interaction, our idea is that these differences reflect reality of interaction. Consequently we do not support correction for gender in the analysis when it is being used in seminars for teams and organizations.

Specific samples (bigger than N=1000)

PARTICIPANTS IN PROJECT MANAGEMENT COURSES AND PROJECT MEMBERS

Table 5. *Project training course participants and project members, N=1010*

DI category	Blue	Red	Green	Age	Men/Fem
SD	7,89	6,39	5,68	8,31	
Mean	33,65	26,19	24,07	39,84	1,37

STUDENTS AT NTNU, TRONDHEIM

This sample N=1491 has been discussed on conferences (AIB conference, 2008) as well as in the book *Diversity Icebreaker – How to Manage Diversity* (Ekelund & Langvik, 2008).

STUDENTS AT BOCCONI BUSINESS SCHOOL, MILANO

Today we have a sample of N=1177, however only a portion of this sample has (N=407) was discussed in conferences as well as well as in the book *Diversity Icebreaker – How to Manage Diversity* (Ekelund & Langvik, 2008).

Norm no 2, N= 8859 (enlarged global norm, 2011)

We often work with groups with mixed cultural background. When it comes to the application of norms in especially in such settings, our focus is not on individual comparison, but on using norms as a guideline for the facilitator and the participants to understand what can "high and low" mean in terms of the Red, Blue and Green in a group during the workshop. For the participants, the global norm – as it presented in the Profile Folder – provides some hints on another reference group, which is different from the local, workshop group and the "here-and-now" processes it is related to, and which can be relevant for self-understanding.

MEANS AND STANDARD DEVIATIONS – APPLICABLE IN NORWAY?

When we created the norm in September 2011, we added 240 data units into original Norwegian sample. We termed this norm the "new Global Norm". These 240 units of data are more diverse

regarding the sampling procedure, which reflects an increasing global use of the Diversity Icebreaker (in terms of scope and areas of application).

Table 6 gives an overview over different sub-groups of data collected for this norm; comments regarding these different sub-groups follow after the table:

Table 6. *The new Global Norm*

	N	Blue	Red	Green	Age	Gender
Global Norm	**8859**	**31**	**28**	**25**[1]		
1. 2006-11 Norwegian samples	2016	30,34	29,19	24,36	40,85	1,61
2. Members in projects/ proj. training seminars	1010	33,65	26,19	24,07	39,84	1,37
3. Students at NTNUniversity, Trondheim 2008	1479	33,20	25,86	25,33	20,26	1,39
4. Bocconi business school students 2008/9/10	1177	30,30	28,89	24,77		
5. First norms from 23rd June 2005	1341	30,61	28,66	24,62		
6. Mixed cross-cultural organizational settings	1012	29,78	26,61	27,54	38,02	1,40
7. Bulgarian business sample	142	26,92	28,47	28,47	34,10	1,67
8. Bulgarian youth school sample age 14-15	377	22,97	32,07	28,88	15,71	1,72
9. Danish samples from business	305	31,37	28,00	24,70	33,07	1,52

Descriptions of the different samples that have been integrated in the new Global Norm:

1. 2006-11 Norwegian samples are the data gathered the same way as the first norm from 2006.

2. Participants in project training seminar and project members: These are participants in project training – or members of project where we as consultants have been trained interpersonal interaction or executed developmental work. Parts of these samples are from globally recruited projects, about 25 % – since an increasing part of our consultation are in globalized project work.

---

[1] Numbers are presented without decimal points for means for the aggregated global sample – this is how we present the norm to the participants in the workshops (e.g. in the *Profile Folder*).

3. Students at NTNU, Trondheim 2008; with the majority being students early in their first years of higher education.

4. Students at the Bocconi business school, year 2008, 2009 and 2010; gathered during the orientation week at the beginning of academic year in autumn.

5. First norms from 23rd June 2005, have been described earlier (as part of Norm 1)

6. Cross-cultural organizational settings – these are samples where clients were international organizations and where the participants also represented a large multicultural background, like multinational companies and MBA programs at IMD.

7. Bulgarian business sample – these data represents similarly recruited samples representing our outreach in the Bulgarian market together with a Bulgarian consultant.

8. Bulgarian youth school sample age 14-15 – data gathered during a state-supported project in Bulgaria, Sofia.

9. Danish business samples- data gathered during our typical consultation work in the Danish market.

As we can see, there are some large samples of students – which also reflects that an important part of our customers are educational institutions (e.g. North Eastern University in USA, Bocconi Business School in Italy, Humbholdt University in Germany and other, post-graduate business schools like INSEAD and IMD). The samples also reflect an increasingly global workforce sourcing, both on the organizational level as well as on the individual level.

One of the reasons for having such a global norm is that the cross-cultural application has been highlighted as one of the important areas of the Diversity Icebreaker application. Furthermore, our attention is growing more and more in direction of a global outreach, both with people, businesses as well as research – and less orientation towards national boundaries with cultural identity as a differentiator.

Table 7 presents the differences in means between first norms (Norm no 1) and the norms created by adding the additional data.

Table 7. *Differences in means between first norm (23rd June 2006) and the added norm*

	Blue	Red	Green
First norm, N=1371 (23rd June 2006)	30,61	28,66	24,62
Added, N=7485	30,91	27,81	25,30
Difference	-0,30	0,85	-0,68

In order to judge whether the differences have practical implications – it is to look upon these differences in the two norms based upon standard deviations in different samples. An overview of different standard deviations imply that they vary between 8,2 and 5,5 (see Table 8. below)

Table 8. *Standard deviations in different samples in the new Global Norm*

St Dev in different samples in global norms	Blue	Red	Green
About 2000 more Norwegian	7,10	6,60	6,80
About 500 first Norwegians	8,20	7,30	6,40
About 1000 international	6,80	6,00	6,20
1010 project participants	7,80	6,40	5,70
1177 Bocconi stud	7,32	6,82	5,69
1340 NTNU students	6,75	6,95	5,51
Simplified presentation for use in the workshops	*7,00*	*6,50*	*6,50)*

Given the standard deviations compared to differences in means between the Norwegian Norm no 1 of N=1378 and the enlarged, new Global Norm, we state that the differences have no significant effect on the practical use of the tool. For this reason we can use the enlarged, new Global Norm also in Norway.

We are aware and make people aware (on the website and in the materials) that we do not use specific criteria either for intra-personal types nor cut-off scores for different groups. Our basic use of the norms is for giving consultants and participants an idea of what is the mean – making it possible to make some kind of comparison with the general population. For this reason we have used this new Global Norm both for Norwegian and other groups globally. This norm is best presented in the Profile Folder, where the standard deviations for each of the scales are taken into consideration when illustrating the distance from the statistical mean.

Sample analysis of global norms: N= 4792, means, gender and age

Descriptive statistics – sample reflecting our typical business customers
In this analysis of N=4792, we have taken away the old norm data from before 2006 norms, student groups from Bocconi and NTNU are also excluded. Thus, the sample reflects the typical business customers of our consultancy.

Table 9. *General descriptive statistics*

	N	Minimum	Maximum	Mean	Std. Deviation
BLUE	4792	,00	66,00	30,9090	7,69979
RED	4792	,00	59,00	27,9978	6,93988
GREEN	4792	,00	63,00	24,9039	6,57635
AGE	3974	,00	84,00	39,4698	12,43170
SEX (M=1, F=2)	3951	,00	47,00	1,5318	1,12910
Valid N (listwise)	3951				

Table 10. *Split by gender descriptive statistics*

SEX (M=1, F=2)		N	Minimum	Maximum	Mean	Std. Deviation
MEN	BLUE	1952	,00	60,00	32,3094	7,85638
	RED	1952	,00	52,00	25,5446	6,74968
	GREEN	1952	,00	63,00	25,9068	6,43323
	AGE	1952	,00	74,00	39,9237	12,58264
	Valid N (listwise)	1952				
FEMALE	BLUE	1995	,00	66,00	29,8784	7,60565

	RED	1995	,00	59,00	30,4175	6,62075
	GREEN	1995	,00	49,00	23,5145	6,78832
	AGE	1995	,00	76,00	39,0266	12,17911
	Valid N (listwise)	1995				

Figure 2. *Means for Red, Blue and Green split by gender – graphical illustration*
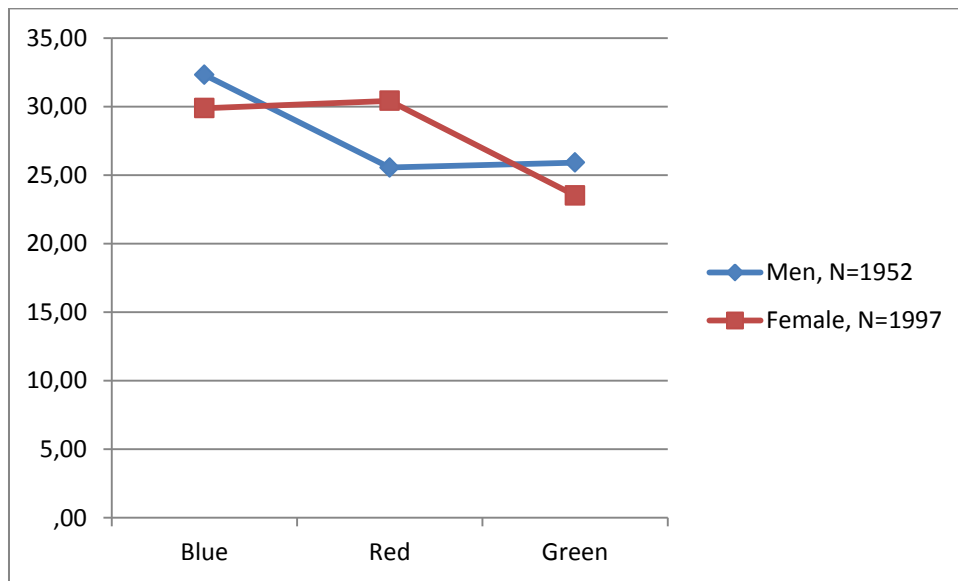
Table 11. *Split by age descriptive statistics* (*N=4792)*

AGE:	<29 N=798	30-39 N=1179	40-49 N=1124	50-59 N=678	>60 N=195
Blue	30,28	31,32	30,75	31,41	33,33
Red	28,83	28,10	27,77	27,67	26,67
Green	24,63	24,35	25,26	24,81	23,75

Figure 3. *Means for Red, Blue and Green split by age – graphical illustration*
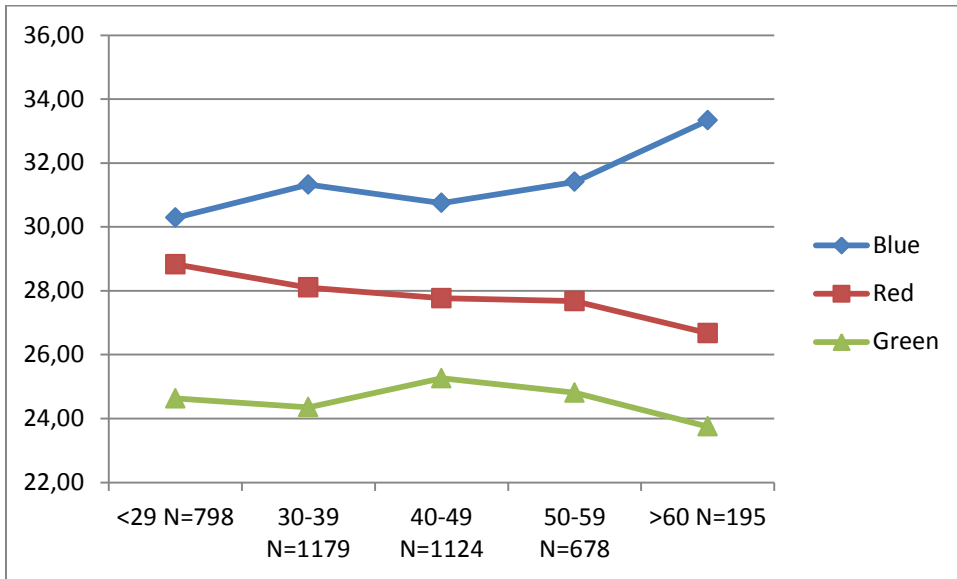
Norms 2 reduced sample of N=4792: Transformation of raw scores

Table 12. *Norms of 10% percentiles N= 4792 – means by percentiles*

		BLUE	RED	GREEN
N	Valid	4792	4792	4792
	Missing	6	6	6
Percentiles	10	21,0000	20,0000	17,0000
	20	25,0000	22,0000	20,0000
	30	27,0000	25,0000	22,0000
	40	29,0000	26,0000	23,0000
	50	31,0000	28,0000	25,0000
	60	33,0000	30,0000	26,0000
	70	35,0000	31,0000	28,0000
	80	37,0000	33,0000	30,0000
	90	40,0000	36,0000	33,0000

Local norms

In practical use we have since 2004 used an Excel matrix for consultants – where they in seminars can punch data of the participants and show means and standard deviation of the current group. It can be compared to the means in norms – or compared to different samples that have been presented and described in the Manual. The intention here is to stimulate collective reflection upon group scores and give an individual a comparison of him/herself to the group. This local creation of comparisons seems to be highly valued by consultants and participants. It seems also to be more in line with ideas of creating a local story – in comparison to a larger population.

In connection with seminars in different organizations and environments we often examine how the whole group scores and set up a group profile. Even in small samples we have seen clear differences that tell us something about how different environments can recruit and nourish different colour profiles. In the figure below some of the samples we use are presented, illustrating some of the differences found.

Figure 4. *Differences between groups*
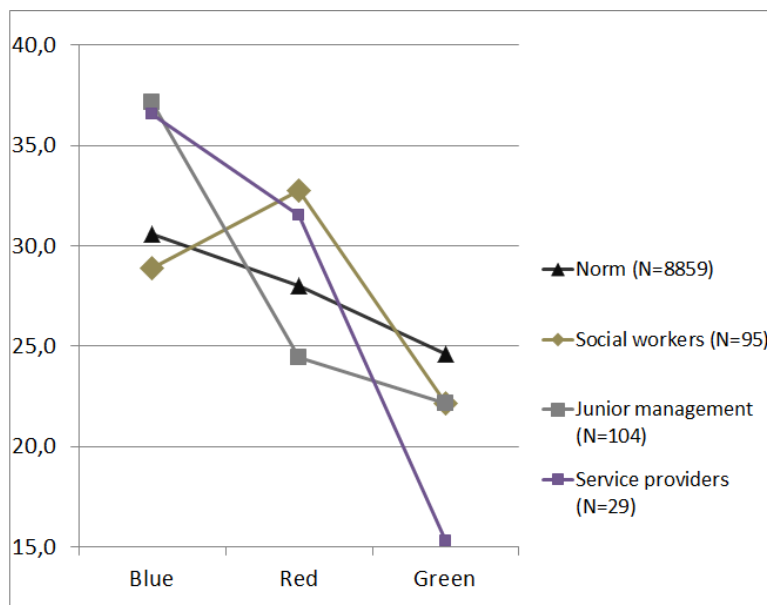
Consultants who buy our products get access to an Excel sheet where they can score the participants results in the seminar (see the *Diversity Icebreaker resources* section above). It can often be used as a tool for collective reflection upon the total results, or sub-sample results between different groups in the rooms (relevant in conflict management or in the development of interaction between departments). When the illustration of the group results compared to other groups are presented in seminars, this strengthens the group's collective perspective of the group as a whole. The participants get to reflect on how they, as a group, differ from other groups.

The examples given in the graph above (Figure 3) illustrate three different contexts that we think are interesting to show some of the differences in scores; one of the samples is role specific (*Junior management*), another is the institutional context, of youth and child care sector, illustrated by the *Social workers*, and the third is the departmental culture of service management in the bureaucracy.

These examples do not represent norms, but local examples that might be used as a way of contextualizing the meaning of the group results.

An illustration of this "contextualization of meaning" is from this article written (in Norwegian) by Britt Rand Hjertnes-Schjødt (2013):

"Deltakere på 3. samling i fordypningsprogrammet i samfunnspsykologi har gjennom flere kull blitt kartlagt med Diversity Icebreaker (DI- Ekelund & Langvik, 2008). DI brukes også i organisasjonsutvikling internasjonalt, og kartlegger kognitive stiler hos enkeltpersoner som funksjonelle i tverrfaglig samarbeid. Tre type kognitive stiler beskrives: I) den planmessige og målrettede blå, 2) den person- og prosess orienterte røde og 3) den entusiastiske, kreative og nytenkende. I et team og i et utviklingsarbeid trenge alle tre typer kognitive stiler for å sikre mangfoldige prosesser og gode resultater. Samfunnspsykologene på fordypningskullene viste seg å ha flere deltagere med mer av den tredje kognitive stilen (entusiastisk, kreativ, nytenkende) sammenlignet med andre kliniske fordypningskull for psykologer og andre helse- og sosial arbeidere (Ekelund, 2011). Dette skulle gjøre samfunnspsykologer særlig anvendelig i utviklingsarbeid. Vi vet ikke om dette handler mest om hvem som søker seg til fordypningsprogrammet, eller om det også er et resultat av påvirkning gjennom fordypningsprosessen. "

Reliability

Reliability of a test is a degree of how precisely the instrument measures a construct (Crocker & Algina, 1986). A measure is said to have high reliability if yields similar results under consistent conditions. However, the operationalization of reliability in the social sciences, where the measured variable is usually latent, is not always easy. In psychometrics, the terms reliability, can be referred to different things or types or reliability: *test-retest reliability, internal consistency, parallel-test reliability, inter-rater reliability,* etc.

This section will present studies supporting the reliability of the Diversity Icebreaker in terms of its internal consistency and test-retest reliability, as well as a study comparing reliabilities yielded by the measure with the partial-ipsative format and the DI questionnaire adapted to the Likert scale. The section will also outline the future, planned studies related to reliability.

Internal consistency

Internal consistency provides the degree to which each item of the instrument is measuring the same construct. The most common way to assess internal consistency is Cronbach's α coefficient, which measures the relation between average inter-item covariance and average variance of items. A value of Cronbach's α>.70 is considered acceptable.

The first reliability study of the Diversity Icebreaker questionnaire was done when the questionnaire was created (Hegge, 1997). The reliability rates measured with Cronbach's α for the Red, Blue and Green categories were from .62 to .70. Improvements in the questionnaire that followed were done partially in order to increase the dimensions' reliabilities.

The version used today is reported to have the following reliabilities: for Red α=.81, for Blue α=.82 and for Green α=75. These reliability coefficients were observed in a study where a group of N=472 respondents filled the test fitted with the partial-ipsative format.

Test-retest reliability

Test-retest reliability (r_{tt}) indicates the degree to which a test yields same scores for each of the respondents when he or she completes the test on another occasion. It is the correlation between the scores of a respondent measured at two different points of time. For instruments which measure individuals' abilities and personality traits, test-retest reliability coefficient should be $r_{tt}>.70$.

The personality traits are relatively stable over time (Soldz & Vaillant, 1999) and although the Red, Blue and Green categories are not personality traits per se, they are influenced by them (see the personality studies part of the *Convergent internal validity* section below), and hence one should expect the scores on the DI dimensions to be at least moderately stable.

To investigate this assumption, a number test-retest reliability study was conducted.

First, the Diversity Icebreaker questionnaire was administered twice to a group of N=56 participants from five different organisations (all of which were in an entrepreneurial stage with a lot of change and learning) The following test-retest reliability coefficients were obtained: for Blue $r_{tt}=.70$, for Red $r_{tt}=.57$ and for Green $r_{tt}=.81$.

These test-retest reliabilities and are considered as moderate to high coefficients. The possible explanation of the lower r_{tt} coefficient for Red could be explained by the relational and context dependant character of this DI category, which makes it more susceptible to change (yet to be tested). However, these results were based on a small sample, and needed to be followed up by a larger-scale study.

A new study of the test-retest reliability was conducted.

Data in that study originated from the following samples:

- Sample 1: N=101; sample consisting of the data set reported above (N=56) and the data gathered exclusively for the new study (N=45). All participants represent different organizations and institutions, time intervals between measurements (T1 and T2) varied from 2 months to 11 months. All participants in this sample took part in Diversity Icebreaker seminars led by Human Factors's consultants. This sample is representative for people that typically take part in management consultancy and training.
- Sample 2: N=126 Norwegian students of entrepreneurship. In this sample there was a two-week interval between T1 and T2; however, in addition, at both times T1 and T2 the participants filled out two versions of the questionnaire – the semi-ipsative and the Likert-scale version; there was a two hour interval between these measurements in different formats (the students attended a non-DI related lecture about management between the measurements). The students received no feedback regarding their results in the context of the Diversity Icebreaker seminar and did not take part in any learning session on the topic.

Data from Sample 1 and 2 were combined for the analysis in three data sets:

- Data set 1: N=101, all cases from Sample 1 included for the analysis.
- Data set 2: this data set is based in Sample 2 – all observations were the participants correctly completed the semi-ipsative version of the questionnaire at T1 and T2 (two-week interval) are included in the analysis, N=91.

- Data set 3: this data set is based in Sample 2; in this data set all the participants that completed both the semi-ipsative and Likert-scale versions correctly twice (with two-week interval) are included (total of 82 participants). However, since all of them completed two different formats of the questionnaire twice, with a two hour interval at each T1 and T2; the number of observations in this study is N=164 and the two-hour time interval is treated as a test-retest interval in this data set.

The test-retest reliability rates obtained in Data set 1, N=101 were satisfactory (for Blue r_{tt}=.872, for Red r_{tt}=.793 and for Green r_{tt}=.838, p<.001) to support the notion that the Diversity Icebreaker measures preferences for communication and interaction consistently across the time.

Also the test-retest reliability rates obtained in Data set 2, N=91 are similar and support this notion (for Blue r_{tt}=.894, for Red r_{tt}=.874 and for Green r_{tt}=.720, p<.001). Since, however, in this data set the criterion of N>100 was not met, these results are only presented to additionally support this notion and point to that the Diversity Icebreaker questionnaire most probably provides a consistent measurement also in contexts different than the organizational context. Samples exceeding N=100 should be used in future studies in that context to confirm this assumption.

Furthermore, also in Data set 3, N=164 – where a two-hour interval was used to create test-retest conditions and each of the participants scored questionnaires fitted with two different response formats (semi-ipsative and Likert) – yielded results supporting that notion (although for one of the scales, Green, it was just below the .7 rule of thumb; for Blue r_{tt}=.834, for Red r_{tt}=.825 and for Green r_{tt}=.655, p<.001).

The lower internal consistency rates for Green in Data set 2 and 3 (Cronbach's alpha α T1=.457 and α T2=.651; and) can perhaps be explained by the group characteristics (entrepreneurship students) and was also reflected in somewhat lower test-retest coefficients for Green in this group, for Data set 1 r_{tt}=.655 and for Data set 2 r_{tt}=.72, p<.001.

These last results, supporting consistency of measurement across two different response formats, suggest that the way in which they answer the questions in the Diversity Icebreaker questionnaire does not influence their results on Red, Blue and Green in a significant way. It may be seen upon as evidence supporting the

Partial-ipsative vs. Likert-scale format reliability

In her presentation at the 11th European Psychology Congress, Eva Langvik (2009) discussed the differences between the use of the Likert scale and the partial-ipsative format. She also mentioned that it is often observed that the ipsative format can inflate the tests reliability (Tenopyr, 1988). She then presented a study where she tested whether this was the case with the Diversity Icebreaker.

In her study, she compared the reliability coefficients for Red, Blue and Green obtained with the use of the original, partial-ipsative format of the Diversity Icebreaker (N=1030, the norm data at that time) with the results obtained using the adapted 7-point Likert scale format of the questionnaire (N=122, NTNU students). The results of the study are presented in the Table 13 below and demonstrate that the reliability coefficients for both response formats are very similar, and thus provide evidence that the reliabilities of DI were not inflated.

Table 13. *Cronbach's α coefficients for Likert and partial-ipsative formats.*

	Blue	**Red**	**Green**
Partial-ipsative	.81	.82	.75
Likert scale format	-.80	.83	.75

Similar assumption is supported by another, unpublished study (SPSS output available) realized in the USA , which indicates even higher reliabilities using Likert scale of 1-100 points obtained in a sample of N=583 (Red α=.869, Blue α=.813 and for Green α=.790).

Moreover, Tetyana Sydorenko in her Master Thesis titled "Evaluating the Validity and Reliability of the Diversity Icebreaker Questionnaire" (2012), also used a Likert scale version of the questionnaire and provided similar results supporting the internal consistency of the Diversity Icebreaker (Red α=.85, Blue α=.84 and Green α=.78, in the Norwegian sample of N=127, civil engineering students from NTNU, M_{age} N/A, 40.2% female).

However, she also observed challenges in replicating equally high reliability coefficients in other samples (German and English) and noted that there might be some Green and Red items that if omitted, would increase the internal consistency of the measure.

Future studies

A large scale, cross-cultural validation studies (we have begun gathering data from 5 different countries and aim at the minimum of N=400 per country, public universities business students, in order to obtain a same age-cohort / profession samples) will also test the reliability of the categories in the Diversity Icebreaker. In the study by Sydorenko the samples were relatively small (N=127 for Norway, N=117 for Germany and N=59 for the English sample) – this larger scale study will help to test the assumption about the malfunctioning Red and Green items, she had made.

We have also begun gathering student data in Norway where we ask the participants to fill out both the standard partial-ipsative and the modified Likert scale DI questionnaire at the same time (aim is N=200).

We hope that these two studies will help to document some qualities of interest regarding the discussion about the statistical qualities concerning the use of partial ipsative format.

Validity

Introduction to the Diversity Icebreaker's validity

The concept of validity is reflected in the question: "Do we measure what we intend to measure?" In the field of psychology, however, the issue of validity is complex and the subject of measurement is a non-tangible and an ambitious one: emotions, cognition, behaviours and often their interaction.

The categories of Red, Blue and Green, representing different preferences for communication, interaction and problem solving, are of this nature as well. In addition, the history of their development and the multi-paradigmatic function they presently have in the workshop increases the

complexity of the concept's validity. During the workshop, participants define their local understanding of Red, Blue and Green in social-construction processes, using questionnaire as a starting-point stimulus, and then drawing upon the discourses unfolding between them in and between the groups.

It means that Red, Blue and Green only partially refer to the established, psychological knowledge created in the traditional scientific knowledge-creation processes.

However, our ambition for the present certification process is to provide an overview of studies supporting the validity of the Diversity Icebreaker within the scope of the classical psychological measure-evaluation paradigm. In the process of doing so, we will to a large extent refer to the concept of "construct validity" in its broad sense (Cronbach & Meehl, 1955; Campbell & Fiske, 1959).

The categories of Red, Blue and Green have not been created by reviewing the literature and referring to other constructs and phenomena described in literature (see the *History of Red, Blue and Green* section above). The model is therefore unique in the way it was created and for this reason we have to give special consideration to build up a meaningful understanding of these categories.

When building the understanding of the three categories with help of the external validation processes, both the historical perspective on the development process of Red, Blue and Green, as well as the practical outcomes of the concept's application are of interest.

For example: The history of development implies to focus more on communicative behaviour and interaction – then on personality. One of the concept's practical outcomes – the effects of a large scale communication and social marketing campaign (described in the *Consequential validity* section) – implies that using these categories as guidance in communicative interaction have at least once been proven to be a cost-effective endeavour. Investigating and supporting the construct validity of the concept is important in this context in order to better understand for what purposes it can be used as well as its limitations.

The classical questions of the internal construct validity are: What is that we find confined within the categories? What is not there? The correlation and regression analysis using different psychological models are of interest for this perspective and the internal validity studies can have both practical and theoretical consequences. One theoretical consequence is the possibility to relate the DI categories to other establish concepts, thus enriching the understanding of Red, Blue and Green and inspiring new research and knowledge creation. This in turn can give us guidelines in terms of the concept's practical application.

Furthermore, there are similarities between the theoretical construction of meaning of constructs within the construct validity tradition and creating of meaning for Red, Blue and Green in the social construction processes taking place the Diversity Icebreaker workshop.

In his thesis about the validation of the "quality of life" construct, Bjørn Z. Ekelund (1983) discussed the similarities between the classical understanding of the construct validity presented by Cronbach & Meehl (1955) and the language theory of Saussure (Culler, 1976). The main conclusion drawn from this comparison is that the *network variance* (manifested in e.g. correlations, regression analyses, group comparisons, etc.) creates the meaning of a construct in the same way as the nomological network defines the meaning of a category (Culler, 1976; Rommetvedt, 1972). The notion of the

nomological network can also be traced in guidelines for a *strong program of construct validation* described later by Benson (1998).

In the DI workshop, the meaning of the constructs emerges in collective social construction processes, where different elements contribute to describing Red, Blue and Green. We believe that the following are the sources influencing the process of creating the meaning of Red, Blue and Green:

a. Content of the questions belonging to each of the categories, i.e. the questionnaire's face validity, is the starting point. During group work, the participants have the second page of the questionnaire with them – where it is indicated which statement refers to which colour – and refer to it when describing the categories.
b. Personal experiences shared and discussed inside the mono-coloured groups and agreed upon as pertinent to the category (this process have similarities to the history of focus groups in 1995 and Moscovici (1984) uses the term "anchoring" to describe the process in which individuals apply the new category in relation to their already established knowledge and history of experiences).
c. Comparison during the presentation stage between the self-description and one made by the others, i.e. comparing the inside and outside perspectives – a process similar to the multi-method strategy (Campbell & Fiske, 1959)
d. This is supplemented by adding a body of research and theory developed around the concept (e.g. norms, correlation results with personality, group comparisons), which adds new reference points for creation of meaning of Red, Blue and Green for the participants. It is made available to the participants in the material (the Profile Folder and/or the Personal Workbook), on our website and via social media, and shared by the consultants.

The three first elements mentioned above (a., b. and c.) are ones beyond the established, theoretical and research knowledge: content of the questionnaire, personal experiences, and group interactions.
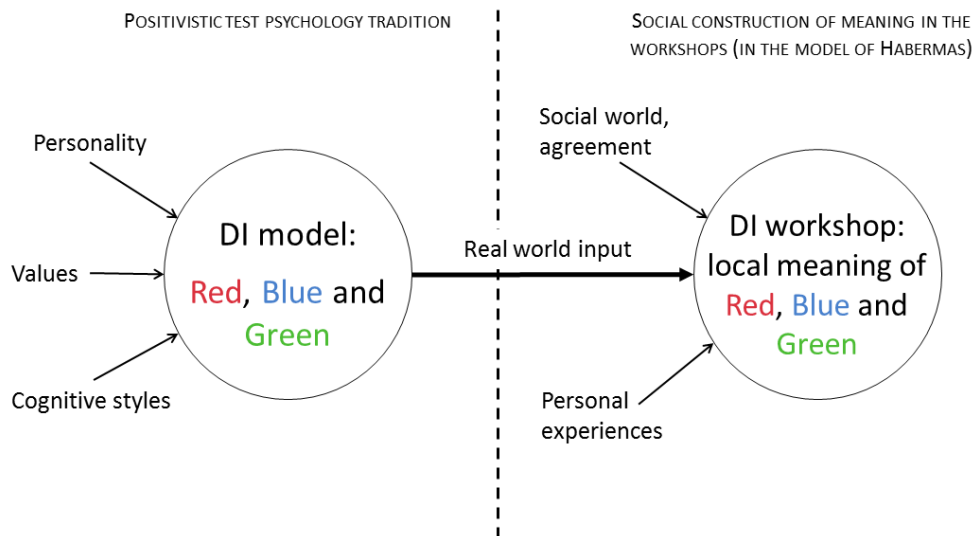
The established research we have built up over the years is thus one out of many stimuli that contribute to the participant's *locally validated* meaning of the categories (point d.).

A paradigm discussion

Thus, we draw a line between the classical, research based knowledge within the psychological paradigm and the social construction of meaning of the categories in practice, in the context-dependant Diversity Icebreaker workshops.

The same type of interaction of different paradigms (*spheres of life*) we find in the work of Habermas on communication (1991), where he refers to the <u>real</u> world, the <u>social</u> world and the <u>personal</u> world as the different sources of meaning. In the DI workshop we state that the "real world input" is what is established in the classical, positivistic test psychology tradition and represented in the questionnaire (backed with research and theory); the "social world" and the "personal world input" are the sources of understanding of Red, Blue and Green in the workshop. Our paradigmatic position is illustrated in the figure on the next page:

Figure 5. *Our view of the model, where the positivistic test psychology tradition and the Habermas theory are integrated to reflect the multi-paradigmatic character of the Diversity Icebreaker (model and workshop)*

POSITIVISTIC TEST PSYCHOLOGY TRADITION

SOCIAL CONSTRUCTION OF MEANING IN THE WORKSHOPS (IN THE MODEL OF HABERMAS)

Personality

Social world, agreement

Values

DI model:
Red, Blue and Green

Real world input

DI workshop: local meaning of Red, Blue and Green

Cognitive styles

Personal experiences

 As a consequence of the multi-paradigmatic understanding intrinsic for the model, which we find highly relevant in our consultative practice and which inspired by the normative model of Habermas; we often state in our work that one of the ambitions of a DI workshop is to create a situation where "people feel free to say what they have in mind and heart, generating social energy that enhances our collective ability to create a better world".

In practice, the Diversity Icebreaker workshop moves beyond the descriptive paradigm of psychology and becomes a developmental, normative model relevant for making a change in interactions between people.
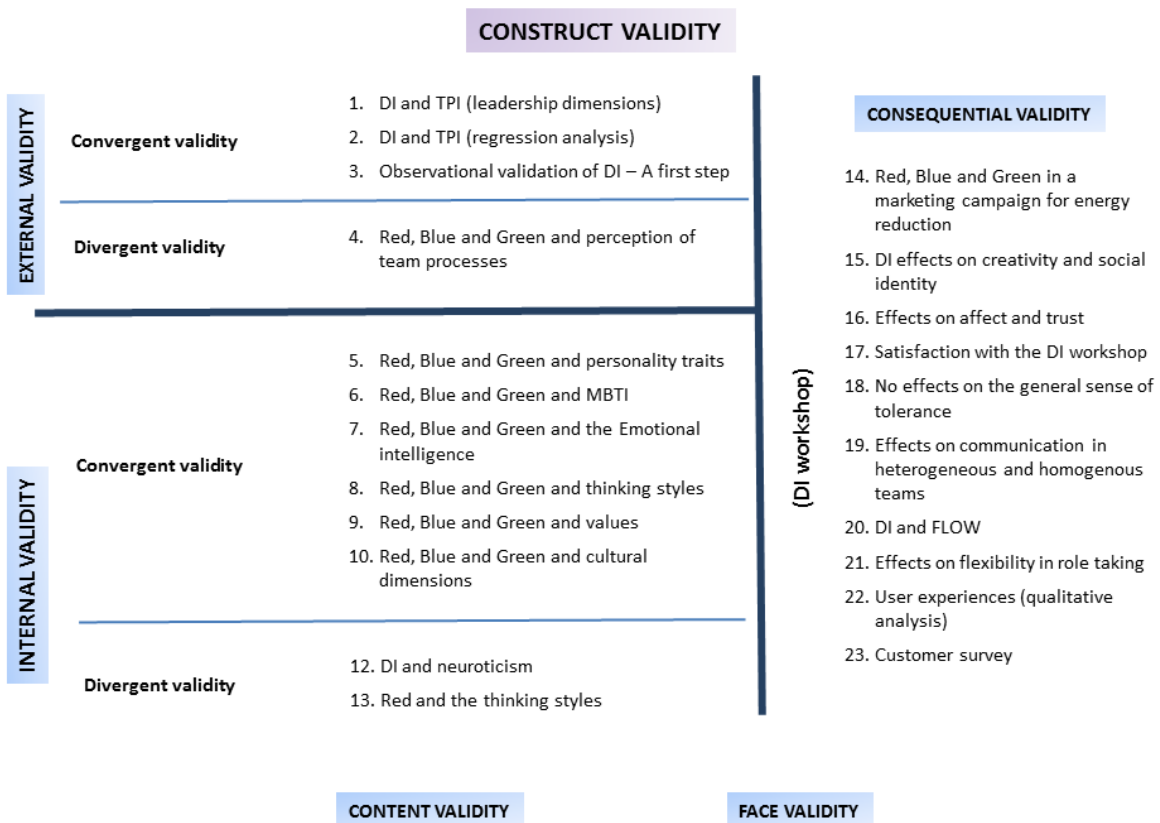
Overview of the section

To provide support for the concept's **construct validity** different **convergent** and **divergent**, both **internal** and **external validity** studies will be discussed (Campbell & Fiske, 1959). Furthermore, other aspects and discussions of construct validity relevant for the Diversity Icebreaker will be also presented: content validity, face validity and consequential validity (Messick, 1995).

Attention is given to the **consequential validity**, with a series of an effect studies presented in that section, because the model's consequences are immanent for its application in the seminar and are what is of the greatest interest for most of the Diversity Icebreaker users.

The body of research and discussions regarding the Diversity Icebreakers validity can be organized as showed on the Figure X (where possible, also indicated are the different studies or themes, which constitute the body of evidence supporting the different areas of the concept's validity):

Figure 6. *The Diversity Icebreaker validity overview*

CONSTRUCT VALIDITY

EXTERNAL VALIDITY

Convergent validity
1. DI and TPI (leadership dimensions)
2. DI and TPI (regression analysis)
3. Observational validation of DI – A first step

Divergent validity
4. Red, Blue and Green and perception of team processes

INTERNAL VALIDITY

Convergent validity
5. Red, Blue and Green and personality traits
6. Red, Blue and Green and MBTI
7. Red, Blue and Green and the Emotional intelligence
8. Red, Blue and Green and thinking styles
9. Red, Blue and Green and values
10. Red, Blue and Green and cultural dimensions

Divergent validity
12. DI and neuroticism
13. Red and the thinking styles

(DI workshop)

CONSEQUENTIAL VALIDITY

14. Red, Blue and Green in a marketing campaign for energy reduction
15. DI effects on creativity and social identity
16. Effects on affect and trust
17. Satisfaction with the DI workshop
18. No effects on the general sense of tolerance
19. Effects on communication in heterogeneous and homogenous teams
20. DI and FLOW
21. Effects on flexibility in role taking
22. User experiences (qualitative analysis)
23. Customer survey

CONTENT VALIDITY   **FACE VALIDITY**

In line with the notion by Benson & Hagtvet (1996), that "numerous studies are needed, utilizing different approaches, different samples and different populations to build a body of evidence that supports or fails to support the validity (...)"; additional purposes of this section are to: a) indicate the areas where the body of research should be improved to better support the tool's validity; and b) point to new, possible areas of investigation, which could also result in better understanding of the concepts and its new, possible applications (e.g. the *On-going research projects* section at the end of this document).

RED, BLUE AND GREEN PREFERENCES FOR COMMUNICATION AND INTERACTION
At this point, and before entering the section of the Diversity Icebreaker construct validity, it is worthwhile providing broader descriptors of the three categories constituting the central elements of the construct in question. However, these descriptions below are only functional descriptions – best we have at the moment, which were often used to formulate hypothesis the studies described below (operationalization). Nonetheless, the Red, Blue and Green categories are perceived more as emerging phenomena, in the seminars but also in terms of their construct validity, as their local meaning and understanding varies to some degree in time and context.

Below is an exemplary description of one of the categories – Blue, together with information about the process of the said description creation. We direct the reader to the *Personal Workbook*, pp. 7-9 for descriptions of the two remaining categories – Red and Green.

Documentation for description and evaluation of the Diversity Icebreaker

Questionnaire itself was the source for creating the first part of the descriptions. This descriptive part was created entirely using the key words from all of the 14 questions, which had bearing for a given preference (colour) characteristics. The example for the Blue category:

People with a strong Blue preference are concerned with being concrete and practical.

They like to calculate and work towards solutions, in a systematic manner. They want things to be useful and serve a purpose. The aim of communication is to solve tasks in a precise way. In decision making processes they want the facts to be presented and they measure the arguments in terms of usefulness and goal achievement. They are concerned with keeping the end result precise and all details correct.

(Ekelund & Rydningen, 2008)

This type of presentation draws upon face and content validity - the degree to which a measure looks like it measures what it was intended to in the eyes of participants – and whether the words refer to key words in the construct to be measured.

The second source was the guidelines that were used in the design of the marketing campaign and the training of consultants in 1994 (see the *History of Red, Blue and Green* section above). Below is an example from the "Blue" guidelines:

- Be down-to-earth, practical, focus on usefulness

- Be logical, goal-oriented

- Use facts and examples

- Focus on details

- Use numbers and calculations

- Be structured and well prepared

(Ekelund & Rydningen, 2008)

The third source was based on the first empirical studies, presented in detail in the sections below: *Internal convergent validity* and *Internal divergent validity*, as well as in the book *Diversity Icebreaker: How to Manage Diversity Processes* (Ekelund & Langvik, 2008). The Blue description as an example of understanding of the categories that resulted from these studies:

Tend to think and consider the consequences before they say or do something concrete. They are good at being focused and goal oriented, with the purpose of completing a task. They are not socially dominant, and do not talk about feelings much or get carried away by the world of imagination. They do not seek excitement for its own sake and are not carried away by torrents of positive emotion. Their everyday life is not characterized by impulsiveness and spontaneous suggestions are mostly seen as disturbance.

(Ekelund & Langvik, Diversity Icebreaker: How to Manage Diversity Processes, 2008)

Still this description is integrated in our "Personal workbook" from 2008 (Ekelund & Rydningen, 2008) that is accessible to the participants. And a rationale for these three different ways of describing and the results are also published in the 2008 book (Ekelund & Langvik, 2008). In all the materials we have the same way of describing Red and Green, too.

Construct validity

As noted before, the understanding of the construct validity in its broad sense, as consisting of both external/internal and convergent/divergent validity (Campbell & Fiske, 1959), together with different add-on elements: content validity, generalizability, consequential validity and face validity (Messick, 1995), will be applied in the present documentation.

EXTERNAL VALIDITY

The external validity refers to a degree to which the results a measure gives can be generalized to other situations and to other people (Aronson, Wilson T, Akert, & Fehr, 2007). In other words, the external validity tells us whether the results show consistency with other situations or measurements beyond the study in question – in line with the multitrait-multimethod comparisons method (Campbell & Fiske, 1959).

This section will report studies building both the convergent and divergent external validity of the Diversity Icebreaker, i.e. present those external studies demonstrating an overlap between DI and other constructs, and those studies demonstrating differences between.

Convergent external validity studies

This section will recount three studies:

Two studies relate the Diversity Icebreaker model to another questionnaire measure – i.e. not an experiment or a real-life measure of the kind often related to external validity studies. However, this measure – the Team Performance Inventory (TPI) – is a measure where all team members report perceptions of how they evaluate team's premises, team processes, leadership behaviour as well as team outcome. It is not an individual assessment (like the Diversity Icebreaker is) – but a team assessment. TPI was created by Michael A. West and Bjørn Z. Ekelund (Ekelund & Jørstad, 2002).

The first study is based upon Kirsten Stuestøl Skottheim "DI in Relation to Team and Leaders Evaluations", presented in the book by Ekelund & Langvik, 2008 and as a Master thesis, for NTNU spring 2008.

The second study has not been published. It shows regression lines between Red, Blue and Green and the different outcome variables in TPI. The results are not significant, with only 21 teams, but the indications are line with face validity - and relevant compared to the more complex interaction models on team roles like Belbin.

The third study is a first attempt on creating a tool to observe "typical Red, Blue and Green behaviours" with subjects taking part in collective problem solving processes, i.e. a way to relate DI to an external, non-questionnaire and measurable reference point.

DI and TPI (leadership dimensions)

In this study, 36 teams and the leaders of these teams answered the Team Performance Inventory (TPI). In addition, the leaders also took the Diversity Icebreaker questionnaire.

Documentation for description and evaluation of the Diversity Icebreaker

In the TPI the leaders and team members evaluated Team Input (Completeness and importance, Autonomy, Interdependency, Team composition), Process (Goal focus, Task focus, Involvement, Conflict management, Learning by experience, Creativity and change), and Output (Goal realization, Satisfaction, Commitment, Innovation, Bridging). Included in the Process there is also an evaluation of the leader skills regarding whether she/he is Social-, Task-, or Change oriented.

From the results of the TPI, the main focus of the study was to see whether it was the leader's qualities that would cause a difference between the team members evaluation and the leader's evaluation of their team. The leader qualities included in the study were gender, age and score on the dimensions of DI (so that the highest results on the DI categories was chosen as a category denominator, e.g. if a leader's highest score is on Blue, than he or she is seen as "Blue" later in the analysis).

There are many studies on leadership, and we could look into whether the leaders' personality would have an effect on the evaluations. This is because results have shown that the personality has an effect on leadership (Hogan & Kaiser, 2005). However, we have used the Diversity Icebreaker, since studies (Langvik, 2006) show a link between DI and the personality dimensions in the Five Factor Model. This is to see whether a leader who is more Social oriented will give evaluations more equal to his/her team, or if a leader who is Task oriented will evaluate a team's results higher than his/her team.

If we take a look at studies on leadership and gender, it is concluded that females are more democratic and participating, than men who are autocratic and task oriented (Trinidad & Normore, 2005; Dennis & Kunkel, 2004). If this is the case, we can pose a research questions if female leaders will give more equal evaluations to their team.

Procedure

The samples were gathered by Human Factors AS. 253 respondents participated in the study distributed on 36 teams. 137 (54%) of the respondents were women and 116 (46%) were men. 19 (52,7%) of the leaders were women and 17 (47,2%) were men. All members of the teams answered the TPI, where as the leader of each team also answered the DI. The team sizes varied from three to thirteen members including the leader, and the average size was 7 members. The mean age of a team varied between 39 and 52 years and the leaders age varied between 31 and 62 years, with an average of 47 years of age.

The variables from the TPI that were used in the analysis were the main factors, Team Composition, Team Process, and Team Results. In addition also the leadership sub factors of Team Process: the Leader as Task oriented, the Leader as Social oriented and the Leader as Change oriented were included.

There was also a composite score indicating the difference between the leader's evaluation and the team's mean evaluation in the TPI. This score was made by subtracting the leader's score from the team's mean score. If this score was negative, it meant that the leader's evaluation were higher than those of the team. Respectively, if the difference score was positive, it meant that the team gave higher evaluations than its leader.

The procedure then was to take a t-test to compare the groups with male versus female leader on their mean scores, and a regression analysis to find out what effect the leaders gender, age and DI score had on the difference in the evaluations between leader and team members.

Results

The results from the t-test showed that teams with male leaders tends to evaluate the Team Process differently than teams with female leaders: $t(34)= (3,105)$, $p<.005$. They also tend to evaluate their leaders differently: Team Leadership: $t(34)=(2,578)$, $p<.05$; Leader as Relation oriented: $t(34)=(2,645)$, $p<.05$; Leader as Change oriented: $t(34)=(2,212)$, $p<.05$. If we take a look at the table below, we see that the female leaders mostly gave similar evaluation of their own leadership as the male leaders do, while the teams with a female leader give their leader higher evaluation than teams with male leaders do. This can explain why the difference score from teams with male leaders are higher and more negative than from teams with female leaders.

Table 14. *Group statistics, differences between teams and leaders score on TPI, with male versus female leader*

TPI factors	Female leader		Male leader	
	Team average	Diff. score	Team average	Diff. score
Team Process (T)	3.772		3.411	
		.087		-.285
Team Process (L)	3.684		3.697	
Leadership (T)	3.639		3.347	
		-.101		-.477
Leadership (L)	3.741		3.824	
Task oriented (T)	3.526		3.352	
		-.115		-.376
Task oriented (L)	3.642		3.726	
Social oriented (T)	3.769		3.429	
		-.125		-.523
Social oriented (L)	3.894		3.952	
Change oriented (T)	3.531		3.243	
		-.139		-.535
Change oriented (L)	3.671		3.779	
Team Results (T)	3.446		3.325	
		.011		-.132
Team Results (L)	3.435		3.457	

(T)= Team evaluation

(L)= Leader evaluation

The results of the regression analysis confirm that the leader's gender is an important factor in the difference between team and leaders evaluation of themselves and each other. On the factor Team

Process, the leader aspects gender, age and the DI together explains 48% of the variance in the team and leaders evaluation of the process. The leader's gender is the only one with significant explanation value: B=-.47. On the factor Team Leadership in the TPI, the same variables explain 30% of the variance between team and leaders evaluation on the factor. Again the leader's gender is the only significant value: B=-.481. On the factors Leader as Social oriented and Leader as Change oriented, the variables explain 27,8% and 32,5% of the variance in team and leader's evaluation respectively, with only the leaders gender as a significant affect. From the factor Team Results, the variables explain 38,7% of the variance between team and leader's evaluation. Here, both leaders' gender (B=-.283) and leader's DI score as Blue (-.033) have significant influence on the variance in evaluations. See Table 15.

Table 15. *Explanation variables in difference score on Team Results*

	B	T	SD
Leaders gender	-.283*	-2.703	.105
Leaders age	-.012	-1.825	.006
Leaders DI Blue	-.033**	-3.922	.009
Leaders DI Red	.009	1.204	.008
R²	.387		

* p<.05

** p<.01

The interesting thing here is that on all regressions, even though they were not significant, the DI Red score always had a positive influence on the difference score between the team and leaders evaluations, while DI Blue always had a negative influence on the difference scores. This means that if the leaders highest score on the DI were Red, the leader would have a slight tendency to give lower or equal evaluation of her/his team, while a leader with the highest score on the DI Blue, will have a tendency to evaluate her-/himself and the team higher than the team members. The results on the leader's gender means that female leaders tend to evaluate their team and themselves equal to or lower than her team, and that male leaders will have a tendency to evaluate themselves and the team higher than the team. If we see this in relation to the results from the t-test, it might not be that the leaders evaluate differently from each other, but rather that the teams evaluations varies to whether they have a male or female leader.

In the regression analysis the DI Green score is not included in the analysis since the initial correlation analysis proved a strong correlation between DI Red and DI Green. Even though it is natural that they correlate since they come from a partially ipsative instrument, it was decided to exclude DI Green from the analysis to avoid an artificially high R² (Johannessen, 2003).
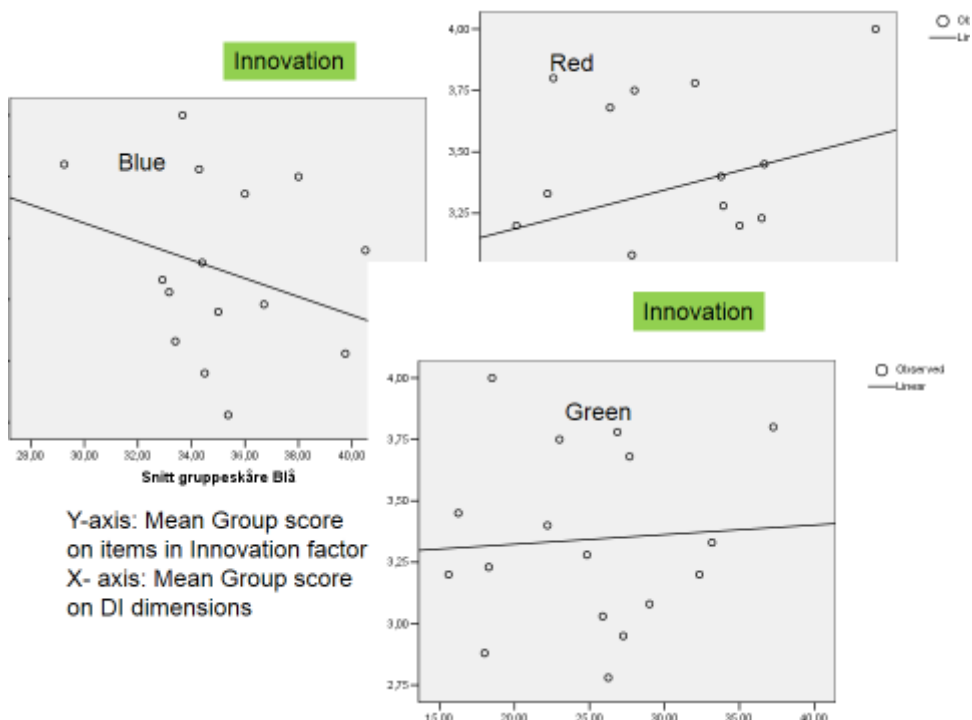
These results show that female leaders are evaluated better than male leaders, and also that teams with female leader evaluate the team higher than teams with a male leader. This must be analysed further, and perhaps tested in organisations. Unfortunately there is little to tell what influence a

leaders score on DI have on the difference in evaluations, perhaps that it is ipsative makes it difficult to use in this analysis, or perhaps it would show better results if we also included the team members DI scores. It would have been interesting to look at the diversity among the team members, when it comes to gender and their DI score. Perhaps the results could show us more about why the female leaders get higher evaluations, and why teams with female leaders tends to evaluate themselves higher than teams with male leaders. For further reading about the study, see Skottheim (2008).

DI and TPI (regression analysis)

Since Belbin's book on management teams it has been a quite normal assumption that the variety and interaction between team participants is essential in order to create a high performing team managing complex tasks. And the ideas of Diversity Icebreaker used as a team role concept is built upon this idea, too (Ekelund & Jørstad, 2002). Anyhow it seems to be quite difficult to document that you need all colours to succeed. There are so many different alternative ways of describing optimal balance between different types of people – count highest scores? Look upon alternative roles for each individuals? What about the number of participants that are quite close to mean – are they flexible or do they add unique qualities etc. In 2005 Bjørn Z. Ekelund took part in a PhD training on hierarchical linear modelling program in Denmark, and based upon having data for all team members in 21 team (about 140 members) both on DI and TPI – he started to look for interaction models that could in a meaningful way predict for example «innovation». The sample of teams and individual scores was gathered as a part of a project where Human Factors contributed to a total team reorganization of the Urban Planning Office of one of the major cities in Norway. It was a project that lasted from 2003 to 2006. It the analysis Bjørn Z Ekelund found no interaction model that created any kind of significant result on any of the TPI outcome variables. The only meaningful pattern, that fits into a more simple model like e.g. «If you have more Blue preference in the group – there will be more Task results achieved.». The regression line between the 5 outcomes variables look like this (Figure 6):

Figure 7. *Regression analysis TPI Innovation and Red, Blue and Green*

Y-axis: Mean Group score on items in Innovation factor
X- axis: Mean Group score on DI dimensions

Documentation for description and evaluation of the Diversity Icebreaker

There are some interesting ideas emerging from these data.

Red – seems to have a positive contribution in all directions.

Blue and Green seems to act in line with what could be expected due to common sense.

Little consistency in direction in relation to Bridging – which also make sense, since no theoretical face value connection could link Red, Blue or Green to acting in a way that did not create problems for other teams in the organization.

Observational validation of DI – A first step
Yet another study pertaining to the area of the external convergent validity of the Diversity Icebreaker is one designed and conducted by Felix Block of the Friedrich-Schiller-University in Jena, Germany. It was a pilot study (the author conducted while he was on an exchange project at the University of Life Sciences, Ås, Norway) attempting to provide a preliminary answer to the following question: "whether or to what extent the [DI] questionnaire results of the participants are reflected in a measurable behavioural outcome" (Block, 2012).

The reader should note that it was only a pilot study (with a small and non-representative sample and new-developed measurement methods used), which implies that the results have a limited bearing for making valuable inferences about the concepts validity. However, we stress the extreme usefulness of the study as to showing new possible directions of testing the Diversity Icebreaker's external validity and developing a "passable and extendable observation scheme" for recording, coding and analysing quantitatively behaviours that can be theoretically related to the Red, Blue and Green categories.

The study
The study was to test if there will be consistency between the participant's results on Red, Blue and Green – obtained in the Diversity Icebreaker questionnaire – and observable, behavioural facets theorized as indicators of either of the DI categories. These facets were operationalized as follows:
   a) *use of pronouns* – the idea was to contrast independence related to the Green category (frequent use of first-person, singular pronouns, e.g. *I*, *me*) with interdependedness and group devotion related to Red and partially to Blue (frequent use of first-person, plural pronouns; *we, us,* etc.).
   b) *affiliative interaction* – this type of interaction drew on a concept developed by Leeder et al. (under review), describing a type of interaction resulting in high sense of belonging and security among group members. A categorization system was proposed to record quantitatively manifestations of affiliative interaction. It was expected that there will be a positive relationship between Red and the affiliative interaction**.**
   c) *problem solving* – the theory of the problem solving process by Albers et al. (2005), with different conceptual modules (e.g. *Situation Analysis, Problem Containment, Search for Alternative solutions*), was applied to structure the observation and test the assumptions of relationship between the models and the Blue and Green categories. Green was expected to be positively related with the modules related with generating and discussing new solutions, whereas Blue was expected to be especially related with those related to implementing of the solutions. Instances of a subjects opening new module (phase) and following into an already opened module, and pertaining to either of the modules, were recorded.

    d) *"within" vs. "beyond" time and space orientation* – this category was finally not measured in the study, as it proved itself not relevant to the nature of the practical task presented to the participants.

Results

The first area of observation - a) *use of pronouns* – yielded no significant correlations between either of the DI categories and particular pronouns.

The second area of observation – b) *affiliative interaction* – yielded results surprisingly consistent with the expectations. Indeed, the higher subject's result on Red was, the more instances of *affiliative interaction* in his/hers utterances. This indicated that there is a strong relationship between the Red category and the concept of affiliative interaction.

The third area of observation – c) *problem solving* – did not produce clear correlation between any of the DI categories and either of the problem solving modes.

Discussion

Both, the results that seem to support the concepts validity in this study (the observed consistency between the Red results and frequency of the *affiliative interaction*) as well as the more ambiguous or non-significant results (no relation observed between the results on Green and Blue, and utterances/behaviours falling to either of the problem solving modules categories), should be interpreted with caution. The study was performed on a small sample (N=8) and many of the participants obtained rather equilibrated scores on the DI categories, which could have resulted in a weak observable effect of their preferences on either of the theorized, Red, Blue or Green behavioural schemes. Furthermore, author notes that the group task chosen for the experiment could have been too restrictive as to allow for some of the investigated behavioural patterns to emerge. He also notes that in terms of the observation area/research question a), the method was imperfect as to distinguishing between different contexts in which the pronouns had appeared and which in turn could have bearing for their Green, Blue or Red character.

However, the very clear pattern revealed between Red and the afiliative interaction, where the results on the first seemed to have a clear influence on the latter, is interesting; and – if replicated – could support the external convergent validity of the Diversity Icebreaker.

More importantly still, the study has revealed and blazed the trail for a new perspective on research related to the Diversity Icebreaker's external validity of a huge value and provided us with a tool for gathering quantitative observational data, which – although it requires further refinement – can be applied in similar, experimental settings.

Observational validation – the next step

The present study was conducted in order to supplement the documentation for description and evaluation of the Diversity Icebreaker – a psychological questionnaire – as part of the DNV psychological tests certification process. The aim of this study was to investigate the concurrent validity of the model in relation to a "real-world" criterion

Introduction

The study described here was conducted in order to supplement the documentation for description and evaluation of the Diversity Icebreaker, which was submitted for of the DNV psychological tests certification process (compliant with the EFPA review model version 3.3) in June, 2013.

In their feedback, the evaluators pointed out two areas for improvement – one of which was the lack of a satisfactory concurrent validity study that would validate the categories of Red, Blue and Green in relation to a real world criterion. The memo from evaluators pointed out that:

Most of the studies presented in the Diversity Icebreaker documentation use other instrument scores and not real world criterion measures. As far as the reviewers can see, only a minor pilot study (Block, 2012 - above) uses such a measure.

A study where a real world criterion measure is correlated with the three colour categories/scales should be carried out.

The EFPA review model defines the concurrent validity studies as follows (Bartman, Lindley, & Kennedy, 2004, p.22):

> Concurrent validity (…) refers to studies where real-world criterion measures (i.e. not other instrument scores) have been correlated with scales.

In terms of the Diversity Icebreaker and the Red, Blue and Green model it seemed most adequate and natural to choose a real-world criterion that would be a) closely related to the theoretical background of the concept and b) reflecting the way the model is most typically used in the real world (i.e. in context of improving communication and interaction quality of people working together). In addition, it was also deemed adequate to intent to expand the scope and precision of the pilot study by Block (2012).

We have chosen the way people interact with each other in a real-world situation, e.g. group decision making, as the variable to be operationalized and measured in the present, concurrent validity study of the Diversity Icebreaker.

We have designed a study where independent judges would evaluate individuals interacting with each other in terms of prevalence of Red, Blue or Green behavioural, using a specially designed observation form. The results from the form would be then tested against the DI questionnaire results of the individuals being observed, to determine coherence – or lack thereof – between the test scores and observable behaviour.

Research material development: the video

It was decided that using recording of people interacting with each other when problem solving and decision making would be most convenient in terms of the ease of application and in order to guarantee the similar study conditions.

Four individuals, who work and interact with each other on regular basis and have developed certain communicational patterns, were recorded when discussing two different topics (either planning of the next Christmas party or searching for a new place for an office to rent). Both the usual, "office" topics as well as the fact that the individuals interact together on regular basis, was seen as beneficial to the objective of using a real-world situation as material in the study.

One recording was chosen for the study (Christmas party), due to the provisional qualitative analysis of its content, made by one of the researchers, which revealed more Red, Blue and Green behaviours relevant for the study.

Research tool development: the observation-form

It was decided that a behaviour observation-form would be developed and later used by the judges in the study, as a way to structure and quantify their observations of the Red, Blue and Green behaviour".

Development of the observation-form involved three stages:

- First, - and in order to improve the quality of this work compared to Felix Block's pilot study - a pool of behavioural descriptions of Red, Blue and Green, based in the concept's theoretical background (primarily (Yukl, 2001; Ekvall & Arvonen, 1991; and Schwartz 1992) and two advanced Diversity Icebreaker users' practical experiences with applying the tool in management consulting, was developed with a prerequisite of being "easily observable". The initial pool consisted of 86 behaviours (28 Blue, 29 Red, 29 Greens; see Appendix 1).
- Second, that pool was sent to a group of experienced Diversity Icebreaker users (N=550, consultants and HR-specialists, experienced in practical application and having a theoretical understanding of the Red, Blue and Green categories) in form of a survey. In the survey, the respondents were asked to evaluate each of the behaviours as Red, Blue or/and Green (a multiple answer format was employed, allowing assigning one, two, three or none of the colours to a behaviour; the behaviours were presented in a random order).
- N=51 respondents filled the survey and based on their answers, the initial pool of behaviours was reduced to 33: the "purest" 11 behaviours per colour were left (i.e. behaviours which were most unanimously indicated as either Red, Blue or Green; the following cut-off scores were used: in case of the Red category – 94% or more of the answers indicated a behaviour as "Red"; in case of Blue– 88% or more; in terms of Green – 72%; see Appendices 2 and 3).

Lastly, an observation-form was created, where the behaviours where listed by colour on an A3-size sheet of paper in order to facilitate the observation process for the judges (see Appendix 4 – attached PDF file). The response format was chosen where the judges would be asked to put a "tick" next to a person and behaviour each time they observe this behaviour with this person.

Documentation for description and evaluation of the Diversity Icebreaker

Note: the initial pool of 86 behaviours and the observational-form were created independently and after the recording of the video-material was made, i.e. the observation-form bore no influence on which behaviours were displayed by the individuals interacting in the video; and vice-versa – the video's content has not influenced the creating of the observation form.

Study

Samples

There were two groups participating in the study.

- Sample 1 consisted of N=12 users of the Diversity Icebreaker concept familiar with the Red, Blue and Green categories (note: none of them participated in the survey described in point b), section above). This sample provided N=48 observations (12 judges x 4 protagonists).
- Sample 2 consisted of N=38 psychology students, who became familiarized with the Red, Blue and Green categories before the study by participating in a classic Diversity Icebreaker workshop. This sample provided N=152 observations (38 judges x 4 protagonists).

Procedure

The participants were first presented with the task and given the time to familiarize themselves with the observation-form. Then, the 8 minute video was played and stopped every 30 seconds for between 30 seconds to 1 minute (depending on the group's readiness to continue) in order to give the participants the time to indicate how many times each person exhibited a given behaviour listed on the form.

After the video was over, the participants were given time to go through the observation-form again and correct it. The completed forms were then collected.

In addition, in case of the Sample 1 (the DI users) the participants were also asked to share feedback with the researchers regarding the study and observation-form design. (Comments most relevant for the present study's purposes are included in the discussion.)

Results

The filled observation forms were coded by counting all the ticks (indicating how many time a given behaviour was exhibited by person 1, 2, 3 and 4 starring in the video). Scores were per item per person were entered for the purpose of reliability analysis and composite results for Red, Blue and Green were calculated for later correlation analysis. There were 4 protagonists in the video and a total of N=50 participants (judges) and a total of N=200 observations (4x50) was collected.

Before presenting the principal results of the study, i.e. correlations between the observed Red, Blue and Green behaviour scores and the Diversity Icebreaker questionnaire scores of the protagonists; the reliability of the observation form as well as the inter-rater reliability of the input provided by the participants (judges) is discussed.

Reliability of the observation-form

Cronbach's alpha internal reliability coefficients were calculated for the complete data set for the whole observation form and for each of its subscales (i.e. Red, Blue and Green composite scores). Furthermore, the reliability coefficients were also calculated similarly for each of the two sub-samples (Group 1 and Group 2) in order to determine possible difference in the way experienced DI-users and students scored the form.

Following are the reliability coefficients obtained for the complete data set (Sample 1 and 2): overall observation-form reliability α=.93; Red scale reliability α=.844, Blue α=.818, and Green α=.817.

Following are the reliability coefficients obtained for the DI-users sub-sample (Sample 1): overall observation-form reliability α=.953; Red scale reliability α=.881, Blue α=.875, and Green α=.881.

Following are the reliability coefficients obtained for the students sub-sample (Sample 2): overall observation-form reliability α=.875; Red scale reliability α=.633, Blue α=.735, and Green α=.633.

Inter-rater reliability between the judges

In this analysis the inter-rater reliabilities were obtained by calculating the Cronbach's alpha coefficients in a data set where the judges were treated as variables. Following are the inter-rated reliability coefficients for the complete data set (Sample 1 and 2) α=.990; the DI-users sub-sample (Sample 1) α=.944, and the students sub-sample (Sample 2) α=.990.

Correlations: behaviour and the DI questionnaire results

Pearson's two-tail correlation was used to determine to what extent the Red, Blue and Green results from the observation form, assigned to the protagonists in the video-recording, are or are not coherent with the results from the Diversity Icebreaker questionnaire these same protagonists obtained. Pearson's *r* was calculated for composite variables, i.e. for total Red, Blue and Green scores from observation-form and total Red, Blue and Green results from the DI questionnaire.

Tables 16, 17 and 18 present the complete correlation matrices for consecutively for: the whole data set, Sample 1 and Sample 2:

Table 16

Correlations between the Red, Blue and Green behaviours – B (observation form) and the test scores – T (the Diversity Icebreaker questionnaire); complete data set, N=200 observations.

	Red B	Green B	Blue T	Red T	Green T
Blue B	-.222**	-.304**	**.809**	-.503**	-.046
Red B		.487**	-.378**	**.508**	-.285**
Green B			-.507**	.404**	**-.070**
Blue T				-.494**	-.202**
Red T					-.752**

*Note. *p<.01, **p<.001*

Table 17

Correlations between the Red, Blue and Green behaviours – B (observation form) and the test scores – T (the Diversity Icebreaker questionnaire); Sample 1 data set, N=48 observations.

	Red B	Green B	Blue T	Red T	Green T
Blue B	-.208	-.296*	**.743****	-.455**	-.050
Red B		.576**	-.470**	**.509****	-.216
Green B			-.560**	.422**	**-.051**
Blue T				-.494**	-.202
Red T					-.752**

Note. *p<.01, **p<.001

Table 18

Correlations between the Red, Blue and Green behaviours – B (observation form) and the test scores – T (the Diversity Icebreaker questionnaire); Sample 2 data set, N=152 observations.

	Red B	Green B	Blue T	Red T	Green T
Blue B	-.344**	-.376**	**.858****	-.537**	-.046
Red B		.375**	-.376**	**.564****	-.350**
Green B			-.508**	.416**	**-.083**
Blue T				-.494**	-.202*
Red T					-.752**

Note. *p<.01, **p<.001

Furthermore, the relationship between the DI questionnaire scores and observational-form scores was also investigated when the latter where "standardized" (as % of 84, in order to provisionally reiterate the original DI questionnaire format). Table 19 presents these results for the complete data set:

Table 19

Correlations between the Red, Blue and Green behaviours – B (observation form) and the test scores – T (the Diversity Icebreaker questionnaire) – "standardized" scores; complete data set, N=200 observations.

	Red B	Green B	Blue T	Red T	Green T
Blue B	-.594**	-.769**	.796**	-.595**	.066
Red B		.014	-.435**	.496**	-.229*
Green B			-.657**	.372**	.079
Blue T				.110	.217**
Red T					-.134

Note. *p<.01, **p<.001

<u>Discussion</u>

Correlation results

We have no knowledge of studies of same character where personal preferences of similar type as in DI have been correlated with observational criteria. However, a meta-analysis of different self-other correlation study by Harris and Schaubroeck (1988) revealed that a mean correlation between self-report and the evaluations oscillated around r=.35. This is explained by the fact that individual behaviour is not only determined by personal preferences, but also by one's role, other people in the interaction, task and context. The Diversity Icebreaker questionnaire is designed to measure personal preferences.

Therefore, the results in our study1 have to be looked upon as more than satisfactory with the mean of the correlations of Mr=.46. However, the differences between the significant and very strong Blue dimension results (.809**) and the moderate high Red results (.508**) vs. the very low and insignificant correlation between Green questionnaire and the observation form (-.070) need to be addressed.

1 Only the raw results for the whole sample N=200 observations are referred to in this section of the discussion.

What could be the possible reasons for the low Green dimension correlations?

Video material

As to avoid any influence between the way people in the video acted and the development of the observation form, the video was developed without a script, guidance and was not steered in any other way to trigger display of all the features of behaviours represented in the DI categories of Red, Blue and Green. The video content was captured independently of the DI categories and even before the observational form was created.

Therefore, it is possible that the video does not cover the empirical domain of the observational form completely, because it was meant to depict more natural group-behaviour.

In addition, a qualitative ad-hoc analysis of the video, as well as the feedback gathered among the participants in the first group, points to some issues that may have resulted from the following:

- One person in the video is not speaking much (possible reasons: status differences and Norwegian not being the first language)
- One person primarily Green on the DI questionnaire results, is partially taking the role of a Red conversation leader.
- One person primarily Red, but might have been evaluated as Green because of her eagerness to repeat and elaborate "the Green ideas" of the leader.

Thus, it might be recommended for the next study that either actors playing the Red, Blue and Green "scripts" should be used, as to cover in a more representative way the theoretical and empirical domains of the three categories; or much more video-material should be gathered in a similar way – and in compliance with the initial approach of capturing natural group and interaction processes – and then parts of it should be chosen, edited and balance to better cover the whole empirical domains of the Red, Blue and Green categories.

Furthermore, given that the correlations for Red and Blue were strong and significant, and that the Red and Blue items (behaviours) in the observation form were developed using the same methodology as the Green; there are reasons to believe that the lack of correlations results from the video material and not the measure.

On the other hand, it was for the Green behaviours pool that the cut-off score was the lowest when 11 behaviours were chosen from the initial pool of 29 (i.e. the most of "non-pure" behaviours were used for Green in the observation form; see Appendices 2 and 3).

This could point to that the Green scale on the observation form could be less precise than Red and Blue, but the internal reliability coefficients for all three scales are similar and high: Red $\alpha=.844$, Blue $\alpha=.818$, and Green $\alpha=.8172$, which leads to think otherwise.

2 Reliability coefficients for the complete data-set, N=200 observations.

Observation form

Another notion, supporting the reliability of the observation form developed for the study, is that the significance, magnitude and the direction of the correlations between the two different sub-samples are very similar; the absolute values of differences of correlations between Sample 1 and 2 are following (raw scores): |.055|for Red, |.11| for Blue and |.032| for blue (Green insignificant in both samples). This supports the notion that the observation form provided a consistent measurement across the samples.

Furthermore, the relationship between the DI test-scores and observational form-scores was also investigated when the latter were "standardized" (as % of 84, in order to provisionally reiterate the DI questionnaire format); and again, significant correlations of similar magnitude and direction were

obtained, supporting the notion that the response format had not influenced the analysis results largely. Absolute values of differences (complete data set N=200 observations) are |.012| for Red, |.013| for Blue and |.009| for Green.

Feedback from the expert group (Sample 1)

The participants from the first group that took part in the study were asked to share their impressions and feedback regarding the material, measure and the study design. These are the comments pertinent for the present discussion:

- Some behaviours in the observation form were very concrete and tangible, some required a less unanimous interpretation in the context of the video.
- In the real life, also the negative behaviours and the non-verbal behaviours are pertinent and important for the dynamics of the interaction and personal development. However, the DI categories of Red, Blue and Green are oriented as to emphasize the positive aspects of diversity and purposely leave out the negative behaviours. It was reflected also when the observation form was developed; however, the video also contains the less positive behaviours (e.g. interrupting). However it can be seen as a flaw in the study context, it is in line with the premise behind the Diversity Icebreaker, i.e. focus on working with specific behaviours that can lead to better results.
- Should the observation form be used in practice (of HR-trainings and development) a format where a participant (judge) observes only one other person (protagonist) and not four, would be much more convenient.

Conclusion

The mean correlations between DI questionnaire dimensions and observational data is Mr=.46; however, with some differences between the categories that were discussed above. We view these results, in the context of a concurrent validity study, as satisfactory in supporting the validity of DI. The present observation study is unique in the sense that a separate observation-oriented measure was developed, based on leadership and management theories related to the Red, Blue and Green dimensions, but with different items; whereas typically the judges in similar studies often use same questions as ones answered by the protagonists being observed (e.g. Costa & McCrea, 1988; where same items were used for the observer-form, only phrased in third person). In light of this fact, the correlations obtained for Red and Blue dimensions should be viewed as exceptionally high.

However, the internal differences between the different categories illustrate that there are many different factors influencing behaviour beyond the individual preferences. However, the stimulus material (video) could also have been a reason for this result.

Only minor differences between the two subsamples and high reliability of observation form suggest that it development was satisfactory. In order to gather conclusive research data about the Green category in practical use, another video material – or other real life contexts – is required.

Furthermore, we suggest that in practice a single-person observation form should to be used in training for recognizing/application of Red, Blue and Green behaviour.

Divergent external validity studies

Red, Blue and Green and perception of team processes
In the process of creating Red, Blue and Green in marketing strategy in energy conservation marketing, there were some ideas that people with different colour preferences had a perceptual filter, cognitive style that skewed systematic the way they perceived the world. This model of thinking we find articulated in language – perception discussions (for example Sapir-Whorf and Saussure), in radical constructivist perspective (Wilden, 1976) and in sociology with different ways of living that creates different ways of perceiving, defining challenges and legitimizing solutions (Innvær,1999). We have measured individuals with 2 different team perception tools, CC (Cross Professional Checklist) and TPI + Leadership feedback. Both tools ask the individual in a questionnaire to report how they have experienced team processes. We found no significant correlations at all with 70 persons from 4 different seminars with CC.

On TPI + Leadership feedback with 135 persons we found only one significant correlation, which we also recognize from our own seminar experiences: The more Green your preference is, the more negative you are towards your own leader (correlation -.23). These results indicate that Red, Blue and Green preferences mostly not influence the way people perceive processes in the team.

INTERNAL VALIDITY

Convergent internal validity

Red, Blue and Green and personality traits
The Diversity Icebreaker model and the Red, Blue and Green categories were not meant to personality traits per se; they were aimed at capturing different preferences in a dynamic setting of interpersonal interactions.

However, there are reasons to believe that the personality influences a person's preferences for communication and interaction to a certain degree, and the research question of investigating the possible relationship between the Diversity Icebreaker and personality is worthwhile to pursue.

The Big Five personality traits or the Five Factor (Costa & McCrae, 1992) has been the most dominant and supported in research model of personality traits. The model consists of five personality dimensions or traits, labelled: Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness, which are assumed to be representing the basic personality traits identifiable both in language and widely supported in other psychological models of personality.

Given the model's predominant position and validity, we decided test the Diversity Icebreaker's validity in terms of personality traits with its help (Langvik, Personality Traits and Team Roles: Introducing a Tricolour Model of Team Roles and its Relationship to Personality Traits in the Five Factor Model, 2006).

Expectations

It was hypothesized that the trait Extroversion (characterized by positive emotions, surgency, and the tendency to seek out stimulation and the company of others) and the trait Agreeableness (tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others) will be positively related to the Red category.

Furthermore, it was hypothesized that there will be a positive relationship between Conscientiousness (tendency to show self-discipline, act dutifully, and aim for achievement against measures or outside expectations) to the Blue category.

The trait Openness to experience (a general appreciation for art, emotion, adventure, unusual ideas, imagination, curiosity, and variety of experience) was expected to be positively related with the Green category.

No expectations were made as to the relation between the trait Neuroticism (the tendency to experience negative emotions, such as anger, anxiety, or depression) and any of the Diversity Icebreaker categories.

Study 1

These results were tested by administering both the Diversity Icebreaker and the NEO-PR (Costa & McCrae, 1992) measures to a group employees in different Norwegian organizations (N=233, M_{age}= 35.64, SD=13.45 and N_{men}=86).

Table 16 below presents correlation analysis results between the five personality traits and the Diversity Icebreaker Red, Blue and Green categories:

Table 20. *Correlation analysis results between NEO-PR and DI*

	Blue	**Red**	**Green**
Neuroticism	-.16	.15	.02
Extraversion	-.35**	.28**	.14**
Openness	-.58**	.14*	.50*
Agreeableness	.06	.21**	-.18
Conscientiousness	.30	-.22	-.12

*p<.05, **p<.01

The correlation analysis manifested a strong, negative relationship between Blue and Openness to experiences and a moderate, negative relationship between Blue and Extraversion; moderate, positive relationships between Red and Extraversion and Agreeableness, a weak relationship between Red and Openness to experiences; there also observed a strong, positive relationship between Green and Openness to experiences and a weak, positive relationship between Green and Extraversion.

Furthermore, multiple regression analyses were conducted in order to estimate the direct effect of the personality traits had on the results in the Red, Blue and Green categories.

Table 21. *Multiple Regression results*

	BLUE		RED		GREEN	
	β	t	B	T	β	t
Neuroticism	-.06	-.85	.15	1.66	-.06	-.74
Extraversion	-.18*	-2.45	.31***	3.39	-.08	-.95
Openness to exp.	-.51***	-6.96	-.03	-.28	.55***	6.68
Agreeableness	-.03	-.39	-.30***	3.81	-.24**	-3.18
Conscientiousness	-.31***	4.19	-.24**	-2.69	-.11	-1.31

Note[1]. *p<.05, **p<.01, ***p<.001.

Note[2]. Blue: $R^{2 \, (adjusted)}$ = .47 (.45), F (5,131) = 23.57, p< .000. Durbin-Watson=1.92.

Red: $R^{2 \, (adjusted)}$ = .22 (.19), F (5,131) = 7.30, p< .000. Durbin-Watson=1.81.

Green: $R^{2 \, (adjusted)}$ = .32 (.30), F (5,131) = 12.37, p< .000. Durbin-Watson=1.86.

Multiple regression analysis showed that personality traits explained 47% of the variance in Blue, main predictors being Openness to experience, Conscientiousness, and Extraversion; 20 percent in Red, main predictors being Extraversion, Agreeableness and Conscientiousness; and 32 percent of the variance in Green, main predictors being Openness to experience, Agreeableness and Conscientiousness.

Study 2

In 2012 a similar analysis that was done in Israel be Lilach Sagiv and her PhD-students group (a part of a bigger research-collaboration project – see the *Consequential validity* section for other studies within this project), where the DI questionnaire and the group was: N=158, undergraduate business students, 47% female; Mage=23; 78% Israeli born, 9% Jewish immigrants, 8% Arabs; the results are very similar in many cases – further supporting the concept's validity and providing preliminary support for that Red, Blue and Green, like Big 5 model, seem to be consistent across cultures (notion that has yet to be tested in further validation studies).

The results of a correlation analysis are presented in Table 18. below:

Table 22. *Correlation analysis results between NEO-PR and DI, N=158 (Israel, 2012)*

	Blue	**Red**	**Green**
Neuroticism	.04	-.05	-.09
Extraversion	-.29**	.19*	.10
Openness to exp.	-.40**	-.05	.52*
Agreeableness	-.32**	.36**	-.09
Conscientiousness	.23**	-.23**	-.02

h*p<.05, **p<.01

(For a complete report from the study refer to "Diversity Icebreaker in the Middle-East: Personality as Predictors of Workshop Implications".)

Discussion

The correlation and regression analysis of the results obtained in the study confirmed the hypotheses and provided proof for supporting the convergent validity of the Diversity Icebreaker, as being partially related to the concept of personality traits.

Red, Blue and Green and Meyers-Briggs Type Indicator
This study investigated relationship between the Red, Blue and Green categories of the Diversity Icebreaker and the Meyers-Briggs Type Indicator (MBTI) types – one of the most popular test used in the corporate and organizational settings in the USA (Gardner & Martinko, 1996). Apart from adding to the body of evidence supporting the convergent validity of the DI, the study also delineated the practical differences between the two measures (refer to Chapter 7 in the book *Diversity Icebreaker – How to Manage Diversity Processes* by Ekelund & Langvik, 2008, for more details).

There are four scales in MBTI measuring: *perception* – consisting of two ends/extremes: sensing (S) vs. intuition (N); *decision making criteria* – thinking (T) vs. feeling (F); *orientation to the outer world* – judging (J) vs. perceiving (P); and *energy orientation* – extraversion (E) vs. introversion (I). These four sets of preferences are combined to form sixteen distinct types.

E-I (extraversion-introversion) measures whether people receive their energy from being with others or from solitude. S-N (sensing-intuition) measures whether people tend to take in details through their five senses or tend to use a more intuitive way of taking in the "bigger picture" information. T-F (thinking-feeling) measures whether people tend to make decisions based on consistent logical rules or depending on the impact on individuals. J-P (judging-perceiving) measures whether people are more structured and like to have decisions made or are more spontaneous and prefer to have options open.

Documentation for description and evaluation of the Diversity Icebreaker

Expectations

Due to the nature of the Blue category in the Diversity Icebreaker (logical, systematic, practical, focused on details, less emotional – see the *Red, Blue and Green preferences for communication and interaction* above) and the descriptions of the MBTI types the following hypothesis was made:

> Hypothesis 1: Higher scores on Blue will be related to higher scores on S in the MBTI S-N scale, higher scores on T in the T-F scale, higher scores on J in the J-P scale, and higher scores on I on the I-E scale.

As it comes to the Red category (focused on people, conversation, solidarity and harmony) and the E-I and T-F MBTI scales, the following hypothesis was made:

> Hypothesis 2: Higher scores on Red will be related to higher scores on E in the MBTI E-I scale and higher scores on F in the T-F scale.

No relationship was expected to occur between Red and S-N and J-P scales.

As it comes to the Green category (big-picture oriented, imaginative, value-based and ambitious) and S-N and J-P scales, the following hypothesis was made:

> Hypothesis 3: Higher scores on Green will be related to higher scores on N in the MBTI S-N scale and higher scores on P in the J-P scale.

Furthermore, some configurations of MBTI-scores on different scales (i.e. the different *types*) have similarities with the Diversity Icebreaker categories, namely:

Idealists (NF) tend to be oriented toward building spirit, authenticity, and meaning. These are the human-oriented values with a future-looking orientation. Therefore, we expect to find significant portions of NFS in groups scoring highest on Red, because this category is related to caring for people and harmony; and also within groups scoring highest on Green, because this dimension is related to making things better. Rationalists (NT) are also oriented toward the future and possibilities, but even more towards building power and competence through science and knowledge. Therefore, we expect to find that groups scoring highest on Green have significant portions of NT.

Guardians (SJ) are oriented towards duty and structure, and are concrete. Therefore Blue was expected to be related to SJ. Finally, Artisans (SP), are oriented towards enjoying the present, they value freedom and equality, which are primarily human oriented values. Therefore, the Red category was expected to be related also to SJ.

In summary, it was expected that the people scoring highest on Blue will tend to fall in the SJ types than other temperaments, the group with the highest Red results to contain more SP and NF types, and Green results to be related to the NT and NF types. The following hypothesis was made:

Hypothesis 4: There will be a relationship between MBTI's temperaments and the DI types[2], such that the overlap of Blue and SJ will be greater than with SP, NT, or NF; greater overlap of Red with NF and SP than with NT or SJ, and of Green with NT and NF than with SJ or SP.

Procedure

Both measures were administered to a sample of N=53 (M_{age}= 27; 23% women) MBA students from a Midwestern University in USA, with a time interval of three months between the measurements (MBTI was administered as first). Fourteen subjects were international students: 4 students from China and 1 from each Iceland, Ethiopia, Nigeria, Taiwan, Canada, Ghana, Thailand, Pakistan, Vietnam and Nigeria; 3 were permanent USA residents originally from Mexico, South Korea, and Ghana (making a total of 2 students with a background from Ghana); the remaining students were both born in and were citizens of the USA (36 students). Seventeen were students of colour; five of these were either permanent residents or domestic, and the remaining 12 in this group were international.

Results

Hypotheses were tested using correlational and regression analysis techniques. Three separate regression models were run: one each for Blue, Red, and Green.

In the present document only a summary of the results is presented – for the complete correlation and regression tables refer to *Diversity Icebreaker – How to Manage Diversity* (Ekelund & Langvik, 2008, pp. 67-69):

Hypothesis 1 was partially supported. The assumption was that higher scores on Blue will be related to higher scores on S in the MBTI's S-N scale, higher scores on T in the T-F scale, higher scores on J in the J-P scale, and higher scores on I the I-E scale. There were observed significant correlations in the expected directions between Blue and I (r=-.33, p>.05), S (r=.41. p>.01), and T (r=.41, p>.01); but not with J. Furthermore, regression analysis demonstrated that Blue significantly predicted I, S and T in the expected directions. These findings support Hypothesis 1 for S on the MBTI S-N scale, T on the T-F scale, and I on the I-E scale, but not for J on the J-P scale.

Hypothesis 2 was supported. The assumption was that higher scores on Red will be related to higher scores on E in the MBTI's E-I scale and higher scores on F in the T-F scale. There were observed significant correlations in the expected directions between Red and both E (r=.45, p>.01) and F (r=-.42. p>.01). Furthermore, regression analysis demonstrated that Red predicted significantly E and F in the expected directions; in addition, it also significantly predicted the P dimension. These findings support Hypothesis 2 for E on the MBTI E-I scale and F on the MBTI T-F scale; however they also indicate a possible relationship with the MBTI J-P scale – a relationship not predicted in the hypothesis.

Hypothesis 3 was partially supported. The assumption was that higher scores on Green will be related to higher scores on N in the MBTI's S-N scale and higher scores on P in the J-P scale. There were observed significant correlations in the expected directions between Green and both N (r=-.67,

---

[2] For this particular study, the subjects were assigned to different Red, Blue and Green "types", i.e. not in line with how the measure is usually applied (as a trait measure, not type); a person was assigned to either of the DI types according to his or her highest results.

p>.01) and P (r=-.34 p>.05). The regression analysis demonstrated that Green predicted significantly N in the expected direction; no significant results were observed for P (this is not surprising given the relatively high correlation between N and P scales in this sample, r=.42, p>.01) These findings support Hypothesis 3 for N in the S-N scale, and partially support Hypothesis 3 for P in the J-P scale.

Hypothesis 4 stated that there will be a relationship between certain MBTI's temperaments and the DI categories, such that the overlap of Blue and SJ will be greater than with SP, NT, or NF; greater overlap of Red with NF and SP than with NT or SJ, and of Green with NT and NF than with SJ or SP. This assumption was tested with. Table 3 contains the frequencies of temperament categories and DI types. Examination of this table indicates that there were only 4 respondents in the red category, consistent with our observations about how this sample differs from test norms as discussed above. Given the small sample size for red, it is nonetheless interesting to note that proportions of each DI type by temperament are approximately as predicted. Blues were 52% SJ, 26% SP, 17% NT, and 4% NF. Reds were 50% SPs, 25% NFs, 25% SJs, and there were no red NTs. Greens were 58% NTs, 23% NFs, 15% SJ, and 4% SPs. These results support parts of hypothesis 4. Specifically, SJs did make up the majority of blues and NTs of green. However, there was a greater percentage of SJs in the red category than expected, which is hard to interpret given the small red sample size, and proportionately NFs were at similar levels in both red and green groups, not just the red type as expected, although six of the eight NFs were also green.

Discussion

Of the twelve possible scale-to-type relationships between the four MBTI scales (E-I, S-N, T-F, and J-P) and the three DI types (Red, Blue and Green), we expected eight relationships to be observed in the study. We found evidence for all but one and, additionally, one unexpected relationship. Specifically, no evidence was found for the hypothesized relationship between J in the MBTI J-P and Blue; and an unexpected relationship between J in the J-P scale and the Red category was observed.

Myers (1998) describes the J type as more structured and past-oriented and P as more spontaneous and future-oriented. The people with Blue preference rely on past data and people with Red preference are present-oriented. Only Green of the DI categories is described as future-oriented. Given the strong correlation between S and J in this sample, it is possible that the need for structured data, characteristic for Blue, was captured entirely by S of the S-N scale; whereas J of the J-P scale captured the non-future-orientation of the Red category. However, further work is needed to see if this finding will be replicated and this assumption supported.

Overall, the majority of findings in the present study were as expected and the results either supported or partially supported the theorized relationships predicted in the hypotheses, supporting the Diversity Icebreaker's convergent validity.

Furthermore, this study also provides partial evidence supporting the divergent construct validity of the Diversity Icebreaker in relation to MBTI: Although the instruments relate to each other in predictable ways, they are not the same, and do not measure the same phenomena or in the same way. For example, however it is likely that most ISTJ and ISTP types in the MBTI, will score high on Blue in DI, most EF types will score high on Red, and most NP types on Green; the other MBTI types, such as ESTP, may score more balanced between two or three of the DI categories, despite pertaining to a single one category in the MBTI.

Red, Blue and Green and Emotional Intelligence

Different components of personality traits are to greater or lesser degree related to the emotional phenomena and aspects of life (Costa & McCrae, 1992) and given the relationship between the DI categories and personality reported above, the question of investigating the relationship of Red, Blue and Green to theories or models of emotions, presented itself as interesting to pursue.

The concept of Emotional Intelligence was chosen to be investigated in relation to the Diversity Icebreaker. Emotional intelligence is defined as: "The ability to monitor one's and others' feelings and emotions, do discriminate among them, and to use this information to guide ones' thinking and action" (Petrides & Furnham, Trait Emotional Intelligence: Psychometric Investigation with Reference to Established Trait Taxonomies., 2001).

Study 1

In the first, exploratory pilot study a measure of Emotional Intelligence called Trait Emotional Intelligence Questionnaire – Short Form (TEIQue-SF) was used. TEIQue-SF measures three dimensions: *Emotional expression* (capable of communicating their feelings to others), *Low impulsivity* (reflective and less likely to give in to their urges), and *Emotional skills* (Furham & Petrides, 2003).

In this pilot study N=31 participants (M_{age}=33,6, SD=12.56, 48% female) answered both measures and the relationship between the TEIQue-SF dimensions and the DI categories was analysed. The results are presented in the Table 19 below.

Table 23. *Correlation analysis results between TEIQue-SF and DI*

TEIQue-SF/DI	Blue	Red	Green
Emotional expression	-.42*	.49*	-.28
Low impulsivity	.52*	.01	-.43*
Emotional skills	.-28	.39*	-.02

*p<.05

Results show that the Red category of DI is positively related to Emotional expression (r=.49, p<.05) and Emotional skills (r=.39, p<.05). The Blue category on the other hand, manifested negative relationships to both of these TEIQue dimensions (r=-.42, p<.05 for Emotional expression and r=-.28, non-sig[3], for Emotional skills), and also a strong positive correlation to the Low impulsivity (r=.52, p<.01). The Green dimension on the other hand was found to be negatively correlated to Low impulsivity (r=-.43, p<.05), indicating that people with a strong Green preference, are probably more impulsive than those with the Red, and – especially – Blue preference.

The results are in line with the theoretical background for and the functional descriptions of the DI categories.

---

[3] The result was not significant, but given the small sample and the exploratory, pilot nature of the study, it is worthwhile to mention.

Documentation for description and evaluation of the Diversity Icebreaker

Similar results were provided in yet another, explorative study, where the Bar-On EQ-i measure of the emotional intelligence was employed together with the Diversity Icebreaker in a sample of N=185 individuals (see Table 20 below).

Table 24. *Correlation analysis results between Bar-On EQ-i and DI*

	Blue	**Red**	**Green**
TOTAL EQ-I	-.17*	.15*	.01
INTRAPERSONAL EQ	-.22**	.15*	.07
INTERPERSONAL EQ	-.17*	.33**	-.16*
Self-awareness	-.26**	.28**	-.04
Assertiveness	-.18*	.15	.05
Independence	-.29 **	.16	.19*
Relations	-.21**	.34**	-.12
Empathy	-.08	.19*	-.15
Social responsibility	.07	.17 *	-.12
Flexibility	-.14	.07	.24**
Happiness	-.12	.24**	-.11

*p<.05, **p<.01

Individuals with higher scores on Red scored significantly higher on Total EQI (r=.15, p<.05), and also significantly higher on the main factor Interpersonal EQ (r=.33, p<.01) as well as on the sub factor Self-awareness (r=.28, p<.01). From this we can conclude that higher Red scores are related to better interpersonal / social emotional intelligence.

Discussion

Both studies provide evidence supporting the convergent validity of the Diversity Icebreaker in the fact that – although they were purely exploratory and no prior hypotheses were made – their results are consistent and both measures (TEIQue and Bar-On EQi) are reported to be compatible (Furham & Petrides, 2003). Furthermore, post-study explanations of the results can be built on the theoretical background and functional descriptions of the DI categories, and especially Red where the results confirm the relational and emotional aspects related to this category.

High scores on Red implies lower scores on Blue and Green (due to the partial-ipsative scoring format, see the *Measurement & Scoring* section), and it is important to bear this in mind when interpreting these results. The observed relationships between DI and the emotional intelligence are consistent across measures and support our understanding of Red, Blue and Green (providing evidence for the convergent validity of the concept); however, the moderate power of the

correlations and lack of significant relationships between the DI categories and some of the emotional intelligence facets also indicates discriminant validity. The correlations between two measures should not be too high in order to support this validity (Murphy & Davidshofer, 2001).

Red, Blue and Green and two thinking styles
In the fall of 2011 we conducted a study investigating relation between Red, Blue and Green with two cognitive dimensions described by Jabri (1992): the connective and sequential thinking styles.

This study was conducted as part of the efforts to investigate the Diversity Icebreaker as a *cognitive diversity model*, following previous suggestions from researchers and practitioners (Ekelund, Rossi, & van Egmond, 2010; Matoba, 2011).

Expectations

The connective thinking style is a preference for considering many factors at once and linking previously unconnected ideas, whereas the sequential thinking style is a preference for following an existing set of logical, sequential routines to resolve a problem (Jabri, 1991). Drawing on the theory and previous research regarding Red, Blue and Green, we had expected positive correlations between Green and the connective, and Blue with the sequential thinking style; and – vice versa – a negative relation between Green and the sequential, and between Blue and the connective thinking style. We had not made any assumptions regarding the relation of Red with either of the styles due to lack of data and plausible theoretical formulations.

Results

We tested our assumptions by administering the Diversity Icebreaker questionnaire together with the measure constructed by Post (2011), to a Norwegian-based sample of N=106. The hypotheses were confirmed, as illustrated in Table 21 below:

Table 25. *Correlations between Red, Blue and Green and the two thinking styles (N=106)*

Dimension	Connective thinking	Sequential thinking
Blue	-.458**	.439**
Red	.065	-.316**
Green	.563**	-.240*

Note. *p<.01, **p<.001

There was a significant and positive correlation between the Blue scale and the sequential thinking scale, r=.439; and between the Green scale and the connective thinking scale, r=.563. Furthermore, the Blue scale was significantly and negatively correlated with the connective thinking scale, r=-.458. Similar result was observed in case of the Green and sequential thinking scales, r=-.348.

The direction of the correlation between Green and the Sequential thinking style was negative, as expected, but we can assume that its magnitude could have been different if not the necessity to

remove one of the items pertaining to the Sequential thinking scale as a result of Exploratory Factor Analysis.

<u>Discussion</u>

The present study confirmed that the concepts of Blue and Green preferences for communication and interaction are closely related to the sequential and connective thinking styles, thus supporting the view of the Diversity Icebreaker as a cognitive diversity model.

Interesting to note, the sum of absolute values of the correlations between Blue and the two thinking styles (.892) and the sum of the correlation between Green and the two thinking styles (.803) were both higher than the sum of correlation values between Red and the two thinking styles (.377). It means that the combination of these two colours explains best the structure represented by the connective and sequential thinking styles. However, correlation between Red and the sequential thinking scale was significant and moderate on its own (-.316). This notion is given more consideration in the *Divergent internal validity* section below.

An on-going research project, using a more advanced measure of cognitive styles – the Thinking Styles Inventory (Zhang & Sternberg, 2006), is aimed to shed more light on this and other questions in effort to look at the Diversity Icebreaker as a cognitive diversity model.

Red, Blue and Green and values

The present study – investigating the possible relationship between the Red, Blue and Green categories and the values model (Schwartz, Universals in the Content and Structure of Values: Theory and Empirical Tests in 20 Countries, 1992) – was a part of a part of a bigger research-collaboration project (see the *Consequential validity* section for other studies within this project).

The Schwartz's value model consists of the following dimensions: *Power* (authority, leadership, dominance), *Achievement* (success, capability, ambition, influence, intelligence, self-respect) *Hedonism* (pleasure, enjoying life), *Stimulation* (daring activities, varied life, exciting life), *Self-direction* (creativity, freedom, independence, curiosity, choosing your own goals), *Universalism* (broadmindedness, wisdom, social justice, equality, a world at peace, a world of beauty, unity with nature, protecting the environment, inner harmony), *Benevolence* (helpfulness, honesty, forgiveness, loyalty, responsibility, friendship), *Tradition* (accepting one's portion in life, humility, devoutness, respect for tradition, moderation), *Conformity* (self-discipline, obedience), *Security* (cleanliness, family security, national security, stability of social order, reciprocation of favours, health, sense of belonging).

A group of N=101 undergraduate business students filled the Diversity Icebreaker questionnaire and a 46-item measure based on Schwartz Value Survey – SVS (Schwartz, 1992). Cronbach's alphas ranged for this measure from .52 to .78.

<u>Results</u>

The researchers correlated participants' score on Red, Blue and Green with their scores on values. Figure 7 below presents the results for the DI categories and values.

Figure 8. *The Red, Blue and Green categories and values*

	sec	con	tr	be	un	sd	st	hed	ach	pow
Red	0,01	0,03	0,19	0,41	0,39	-0,3	-0,2	-0,1	-0,4	-0,4
Blue	0,21	0,29	0,15	-0,2	-0,4	-0,2	-0,2	-0	0,2	0,35
Green	-0,3	-0,2	-0,4	-0,3	-0	0,54	0,47	0,13	0,24	0,12
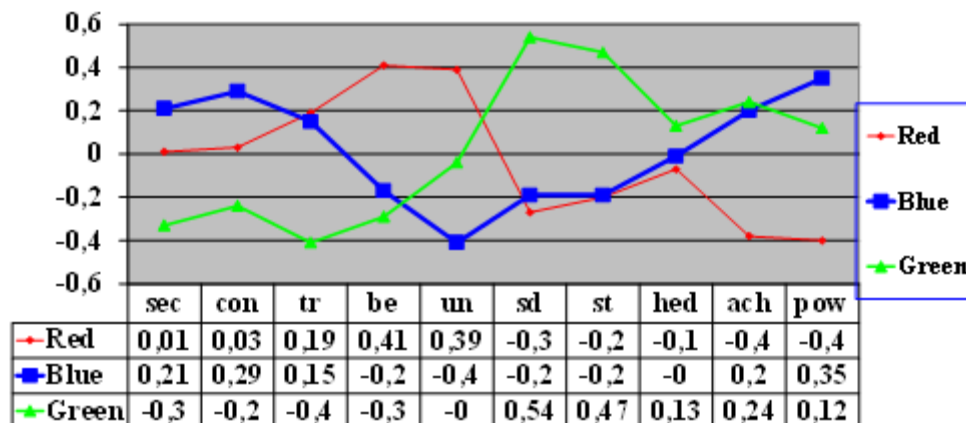
The results show that the three colour categories consistently differed in the motivations (values) that underlie them: Red focuses mainly on good relationships with others, the Blue reflects and emphasizes conservation and power, and the green type emphasizes openness to new ideas and experiences.

(Supplementary regression analyses showed that values and traits had independent effects in explaining the variance in each colour category. Together, values and traits explained 59% of the variance in the red and blue types, and 66% of the variance in the green type. It therefore seems that the colour types represent combinations of traits and values.)

Red, Blue and Green and cultural dimensions
During the years 2006 and 2007 Bjørn Z. Ekelund led personal training of staff from two different engineering and manufacturing companies in five different workshops with participants coming from Norway, Germany, Australia, USA, UK and France. In this setting empirical data was gathered and the possible relationship between the Red, Blue and Green categories and cultural dimensions (Hofstede, 2001) was explored.

Study

The DI questionnaire was administered in the beginning of each of the workshops, later – the participants were also asked to answer the "Cultural Light" questionnaire. The purpose of this questionnaire is to measure individual values related to established concepts from the cultural dimensions (Hall, 1959; Hofstede, 2001; Maznevski, 1994).

Cultural Light is a questionnaire mapping cultural values. It was made first time in 2005 as a tool to create awareness and measure cultural values. The questionnaire has 5 dimensions, but in the context of the present study we will report on three them (which were significant).

The *Individualism* vs. *Collectivism* is inspired by almost all cultural dimensions concepts (Kluckhohn & Strodtbeck, 1961, Hofstede, 2001, etc.). *Performance* vs. *Social* is included in the GLOBE project (House et al., 2004) as well as Hofstede's *Masculinity* vs. *Femininity* dimensios (Hofstede, 2001). The *Monochrone* vs. *Polychrone* dimensions is outlined and inspired by Trompenaars in this setting (Trompenaars,   )

Documentation for description and evaluation of the Diversity Icebreaker

The questionnaire has an ipsative format where different cultural values are positioned as opposite ends of each dimension:

Individualism vs. Collectivism; concern about the individual's vs. collective's best, individual's rights vs. consider the collective, bonus to individual's vs. workgroup, loyal to one's values vs. loyal to the group, individual autonomy vs. group's rules should govern.

Performance vs. Social; goal achievement vs. maintaining relations to others, competition vs. interaction inspires, reward on results vs. positions and education, taking action to succeed vs. reflection and conversation, flexibility vs. rules to be respected.

Monochrone vs. Polychrone: Doing one thing vs. several things at the same time, appointments must be followed vs. not control life, work is separated from privacy vs. blended, stick to schedule is important vs. taking good care of relations, activities should lead to results vs. part of life.

A total of N=188 observations were gathered in the present study (in 14 cross-cultural training seminars in the period 2005-9, about 80% of the respondents are from Norway, 10% from Germany and the rest have varied cultural backgrounds; 24 % are women, M_{age}=39; all the participants worked in oil&gas and metallurgic industries, mainly engineers and researchers by profession). There were significant correlations with three of the dimensions.

Results

The correlation matrix between the DI categories of Red, Blue and Green and the cultural dimensions are presented in Table 22 on the next page.

Table 26. *Correlations between the Diversity Icebreaker categories and the Cultural Light dimensions (N=188)*

	Blue	Red	Green
Individualism	-.13	-.27*	.41**
Performance	.07	-.26*	.19*
Monochrone	.39**	-.25**	-0.20**

Note[1]. * p<.05, ** p<.01

Note[2]. One end of a given dimensions in provided in the table, thus a negative correlation means a positive correlation with the other end of the same dimension.

Evident from the reported results above is that participants with higher Red scores are more Collectivistic than the participants with higher Green scores, who emerge as being more Individualistic. Red is also positively correlated with Egalitarian attitudes in contrast to the Blue dimension, which is related more to the Hierarchical orientation. Red was also found to be more Polychrone than Blue, which is more Monochrone.

Even although the sample is relatively small in this explorative study, there are significant results that indicate meaningful relationships, which could serve as formulate testable hypotheses for future studies. Results of such studies would possibly be useful for tailoring better cross-cultural trainings based applying DI model properly in the group for transferring pedagogical messages. Take for example: If people with Red preferences are more collectivistic, egalitarian and polychrone, how such findings could be utilized in workshops on cross-cultural issues? Can they, for example, constitute an Italian culture, and create norms, rules and shared experiences that are more Italian-like? Will they be better representatives of Italian culture in business games?

Even though there are strong debates on ecological fallacy (Hofstede, 2001; Kirkman, Lowe, & Gibson, 2006) there have been various attempts at making and testing hypotheses across levels. We have put forward a question whether people with a Red dominance could more easily create an Italian culture. What about the other way, will Diversity Icebreaker with Italians come out with a higher level on Red than what is normal?

One way to examine this challenge is based on a database compiled during autumn 2007 which includes data from N=409 business students from Northern Italy. These data are the first large scale data we have gathered abroad. Figure No 2 shows the results compared with an overall norm. And surprisingly enough, results show that there is basically no difference. However, it is worth mentioning that the norm was set based on a large Norwegian sample, N=1371, representing different sectors and total populations of organizational and work life in Norway.

Figure 9. *Aggregated Samples DI Score Comparisons*
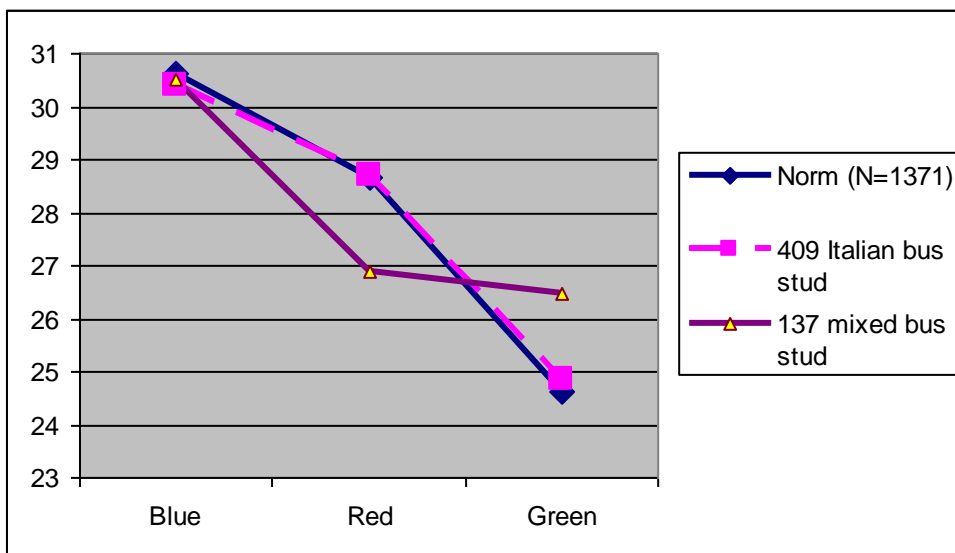
It is also worth mentioning that Norwegians have exhibited significant differences on Diversity Icebreaker type scores between employees in different types of organizations, industries and professions. For example, employees from the social sector score significantly higher on Red. Participants in project management seminars score higher on Blue. Business school students in

Norway score higher on Blue, and less on Red. And creative individuals such as one may find in professional consultants and research & development departments score relatively higher on Green.

As mentioned, the Italian business students had similar scores as the Norwegian norm, but if we compare the Italian results with respondents from same organizational culture, the results are different. We have some preliminary data from undergraduate business students, a combined group of 137 business students from USA, France, Mexico, Denmark and Switzerland. If we make comparisons within the business students segment, Italian business students is the nationality that scores highest on Red and least on Green. The differences between combined group of business students and Italian business students on Red and Green are significant at lower than 1% level. If we combine these data with results on cultural values at individual level described above, we can assume that Italian business students at national level are more Red, implying more polychrone, more collectivistic and more egalitarian in their cultural values.

If we compare the DI scores from the five countries identified in the combined group and the Italian sample and the results with country differences reported on different concepts of cultural dimensions, we could have expanded the construct validity process of the dimensions of Diversity Icebreaker (Cronbach & Meehl, 1955). The early results we have seen are not totally consistent with Hofstede's reports, and more in line with unpublished and updated country score data of Smith & Peterson (sources of guidance), Maznevski (Kluckhohn & Strodtbeck cultural orientations) and Schwartz (Value Systems). These findings serve as a clear invitation for the collection of additional similar data with potential expansions and replications at both levels illustrated here.

The results above tell us that there are probably overlaps between cultural values and preferences in the Red, Blue and Green model measured by Diversity Icebreaker questionnaire. This has implications for the potential of using this instrument in cross-cultural training seminars, which is the main area of concern for this paper. Under these impressions, we now proceed to present what may be the consequences of these recent preliminary findings.

Divergent internal validity

The Diversity Icebreaker is neither just or entirely a personality model or a cognitive thinking styles model. Some of the evidence regarding the divergent validity of the concept had already been mentioned in when some of the convergent validity studies results were interpreted (e.g. studies relating the emotional intelligence to Red, Blue and Green). Additionally, this section will put emphasis on results supporting the convergent validity of DI regarding its relation, or lack of thereof, to neuroticism (part of the abovementioned DI and FFM personality studies) and discussing the Red category relation to the cognitive styles. At the end, a yet not presented study will be discussed, where it was noted that the DI does not influence the perception of team processes.

Neuroticism

In neither of the two abovementioned studies (see the *Convergent internal validity* section), where the Diversity Icebreaker was related to the Five Factor Personality model, was any of the DI categories significantly related to neuroticism.

These replicable result provides support for the divergent validity of the Diversity Icebreaker in the sense that it is not an entirely personality model – not only in terms of how much of the DI results

can the FFM predict (regression analysis above) – but also in terms of not being related to any extent to a whole, one trait in the FFM: Neurotism.

The fact that DI is different by this trait from the Big Five personality traits model has important practical consequences. Neuroticism as a trait is usually perceived as a "negative trait". The Diversity Icebreaker, being deprived of these facets, is thus an attractive tool for opening dialogue and sharing about personal differences – processes that could be impeded by mixing-in the negative characteristics related to neuroticism.

Red and the thinking styles model
In the study where the Diversity Icebreaker was related to the two thinking styles (see the *Convergent internal validity* section above), there were positive correlation results observed between the Blue and Green categories and, consecutively, the sequential and connective thinking styles. There was also a positive and significant correlation observed between Red and the sequential thinking style of r=.316, p>.001.

However, the sums of absolute values of correlations of Blue with the two thinking styles (.892) and of Green with the two thinking styles (.803) were higher than the sum of correlation values between Red and the two thinking styles (.377). It means that it is the Blue and Green preference relate the most and explain the best the two thinking styles model – which is also according to the theory.

Corrine Post (2011) argues that these two thinking styles on the team level have a profound impact on performance and innovation. The study where the Diversity Icebreaker was related to TPI dimensions (see section above *Convergent external validity*) demonstrated that Blue has a negative impact on innovation and Green yielded a positive impact. These results are in line with those relating the two cognitive styles to innovation on the team level (Post, 2011).

However, it was the Red category in the TPI-study which has the strongest, positive impact on innovation. It is possible that Red adds yet another, relational aspect into the cognitive diversity model, which has a strong impact on real outcomes in terms of creativity on team-level.

These assumptions have yet to be tested, but they point to the Diversity Icebreaker as not entirely convergent with the typical cognitive styles division (i.e. to the wholistic and analytical (Jabri, 1991) – reflected in Blue and Green – but adds yet another dimension – Red – with relevance for the cognitive diversity.

Factor structure
Factor analysis can be useful for assessing the concept's validity, because it can answer the question of whether the instrument indeed measures the postulated factors, i.e. whether it captures the structure of the phenomena it purports to measure. Therefore, this method can be applied to especially in assessing the construct validity of the concept.

When it comes to the Diversity Icebreaker, the three categories of preferences for communication and interaction – Red, Blue and Green – were not developed with help of reified scientific methods, like the factor analysis, aimed at obtaining factors that would capture the entire variance of a given variable or facet. Red, Blue and Green emerged in a process of the common sense categorization (Moscovici, 1984) and the questionnaire measuring these categories was developed later by picking questions from other psychological measures by hand, and refined as to obtain the best internal

reliability and face validity of the items per category (see the *History of Red, Blue and Green* section above).

Thus, however it seems natural to think of Red, Blue and Green as factors (e.g. because they form three different scales), by scrutinizing the history of the concept's development, one notices that there is no prior statistical evidence to assume that these categories would perform as such. Therefore, it might be relevant not to think of them as factors, but rather look for another ways of investigating, assessing and refining the internal factor structure of the whole concept. For example, the underlying factors in the DI model could exist across the three colour-categories (meaning that some Red, Blue and Green items could all load on one particular factor).

This ideas has also been put forth by Tetyana Sydorenko in the Master Thesis delivered as partial fulfilment of the requirements for the degree of Master in Business Administration, at the Humboldt-Universität in Berlin titled "Evaluating the Validity and Reliability of the Diversity Icebreaker Questionnaire". She writes:

> "[however] the items belonging to Blue, Red or Green dimension were assumed to measure the same concept [in her study] (…); the items corresponding to one of the colours might cover different aspects, e.g., preferences for communication or working."

These ideas are further explored in Sydorenko's study below.

The DI factor structure
In part of her thesis, Sydorenko performed a meticulous and detailed a meticulous analysis of the factor structure of the model, using a set of statistical methods applicable in factor analysis and not applied previously with the Diversity Icebreaker (EFA and CFA – the latter not discussed in the present document, as it was not performed on the Norwegian sample).

These investigations were to test the assumptions of whether a) the three-colour factor assumption was valid and reflected in the factor analysis, and – if not – what other factor structure could be applicable; and b) whether the factor solution would be stable across different groups.

Procedure and samples

Due to certain limitations as to which statistical procedures can be used with data gathered with partial-ipsative format (the dimensions per se are not independent (Dunlap & Cornwell, 1994), the original questions format was transformed into ordinal scale. Respondents were asked to show their degree of agreement or disagreement on each of the items, so that each item was evaluated independently.

The modified DI questionnaire was administered in three samples used in the study: a) Norwegian (N=127, civil engineering students from NTNU, M_{age} N/A, 40.2% female); b) German (N=117, business administration and economics students from Humboldt-Universität, M_{age}=22.9, 49.6% female), and c) English (N=59, business administration and economics students from Humboldt-Universität, 35.6% female).

EFA for each dimension

An EFA was first performed for each data set (Norwegian, German and English) and on variables representing one colour at the time. The principal component analysis method was used to extract factors and either the Horn or Kaiser criterion was used to delimit the number of factors that should be retained.

For Blue, the EFA produced two- and three-factor solutions (Horn criterion) it was the solution with three factors that was identified as the most plausible and tested in different data sets. They were named: *Preference for working with number*, *Preference for precise communication and decisions* and *Preference for being practical-minded*. 6 items were identified as not belonging to any of the factors and thus a detailed inspection of the item's functioning was advised.

In contrast, the EFA applied for Red category provided a four-factor solution across the data sets (when Kaiser criterion was applied; Horn criterion suggested two factors, but after scrutinizing the items, it was clear that the four-factor solution is most plausible). The factors were: *Preference to be in a group*, *Extraversion, Consideration of other's feelings* (these three factors and corresponding items, appeared to have similar structure in different samples) and *Preference for personal communication* (which seemed to be rather unstable across the data sets).

The interpretation of Exploratory Factor analysis for Green proved itself to be difficult. The amount of extracted factors (based on either Horn or Kaiser criteria) varied from data set to another to certain extent. Furthermore, Sydorenko states that in contrast to Red and Blue, the Green items are difficult to group by content and interpret in light of the EFA, because they deal with different aspects of interaction.

There was one common factor identified (*Positive attitudes towards creative ideas and solutions*), with high loadings of three items measuring positive attitudes towards creative ideas and solutions (these items also had the highest reliability values). However, it was still difficult to offer a joint interpretation of this factor, as there were also other, beside these three, stable items loading, on it in different samples. Also differences in factor loadings between samples for this factor were higher than for Red and Blue.

There were also other factors considered for interpretation, but either due to too many, cross-loading items or due to a too divergent loadings structure, no clear conclusion as to their naming or stability was made. Moreover, there were also a number of items that belonged to different factors in different data sets, which raises the need to scrutinize them.

EFA for the whole set of variables

Exploratory Factor analysis was also performed on the whole set of variables (i.e. on the questions pertaining to Red, Blue and Green all together), separately for each of the sample groups. The purpose was to investigate the relationship between items which belong to different dimensions and possible provide evidence shedding light on whether Red, Blue and Green can be explained by "one-colour factors" (i.e. where items assigned to one colour only load on a given factor).

EFA applied to the whole set of items yielded similar number of factor for each of the samples (the Horn criterion 5-7 and the Kaiser criterion 12-14). Horn criterion was chosen for interpretation, as Kaiser's is deemed not suitable for the analysis with high amount of variables, as it was the case.

A seven-factor solution was chosen for the Norwegian data set and it explained 56% of total variance; all but one factors yielded satisfactory Cronbach's alphas. There were two purely Red factors (*Preference for being in a group and willingness to get to know other people* and *Consideration for other's feelings*), two purely Blue ones (*Preference for working with numbers* and *Preference for practical thinking*), and one purely Green factors (*Preference for imaginative solutions*). In addition, there were two mixed factors: *Preference for precise communication* (predominantly Blue items, however also two Green items with high loadings) and an unnamed factor with three items each from Red, Blue or Green.

In the German sample, there were five factors accounting for 44% of the variance, with only one, purely one-colour factor (consisting of nine Red items). There was one predominantly Blue factor (with a persistent core reflected in the *Preference for working with numbers* notion), but with one strong loading from a Green item. A similar case occurred with an almost entirely Green factor (with one Blue item loading on this factor), which indicates a connection between Blue and Green. Finally, there was also an almost Green factor with a Red item adding to its variance.

There were five factors extracted in the English sample, responsible for 55% of variance, and similar to those obtained in the Norwegian and German samples. Similarly, there was a typically "Red factor" (with two item loadings from Blue and Green categories, which could however be explained by the small sample effect). There was one pure Blue factor as well, but with different items loading on it then in the previous samples (dealing more with the *Preference for practical thinking* than with the *Preference for working with numbers* as was typical for Blue in the previous samples). Finally, there were two mixed factors with high both Green and Blue item factor loadings.

Discussion

The master thesis by Tetyana Sydorenko has important implications for better understanding the structure of the Diversity Icebreaker, assessing the concept's validity and – most importantly – pointing to the areas of future investigation and refinement.

The results indicate that that Red, Blue, and Green should not be considered as separate factors in the tradition, statistical or psychometrical sense – which has been manifested both in the Exploratory and Confirmatory analyses, as well as which is congruent with the concept's history.

Furthermore, the study reports that Red is the most stable and independent of the three categories, whereas Blue and Green were more interrelated. Sydorenko suggests that Green and Blue might represent two, different facets of one dimension – a notion that makes sense when interpreted together with the results of the study discussed previously (in the *Red, Blue and Green and two thinking styles* section above). In this study, both Green and Blue were strongly correlated with the connective and consequential thinking styles consequently, which we know are two opposite facets of one dimension: thinking style. Furthermore, most of the correlation studies with Red, Blue and Green show consistent and negative correlations between Blue and Green, which further supports that that these two categories – or at least parts of them – are interrelated in that way.

One also has to remember about the limitations of this study and the abovementioned conclusions, when referring to the results:

a) The groups in the study was relatively small (N=59, N=117 and N=127) and often bigger samples are required to properly investigate measures factor structure (Costello & Osborne, 2005).

b) The groups had different cultural backgrounds, which, in a sense, made the described study also a cross-cultural validation study. That means that some of the conclusions made by Sydorenko refer to the notion of "whether the factors structure is stable across the cultures", and not whether it is stable by its own, e.g. in the Norwegian population. Future studies may focus on first investigating and refining the factor structure within one culture first, e.g. in a series of test-retest studies, before conducting a cross-cultural validation.

c) The study used the modified Likert scale version of the questionnaire and although previously no significant differences were observed in terms of internal consistencies between these two response formats (see the *Partial-ipsative vs. Likert scale format reliability* section above) the item-loadings and the degree of the factor-overlap could be very different in the two formats. Conclusions from this study are thus to a limited extent applicable across the formats, i.e. to the latent traits underlying the factors in the Diversity Icebreaker.

d) In her study, Sydorenko applied the Principal Components extraction method for the Exploratory Factor analysis. This method is often used, however some authors suggest that the Maximum Likelihood extraction method is better for social and behavioural sciences, where the measurement is characterized by a certain amount of error (Costello & Osborne, 2005). The PCA method takes all variance into analysis into account when extracting factors, whereas the Maximum Likelihood does not take the variance of the error of measurement into account.

In the future, it will be worthwhile to better understand and possible obtain a clear and stable factor solution for the Diversity Icebreaker questionnaire. It will be useful for cross-cultural validation purposes, better understanding of the construct validity and procuring short forms of the test.

In order to do so, we could pursue the notion of different factors that exist across the colours. Hypotheses should be made related to the theoretical factor structure in the Diversity Icebreaker, and could be based on the validity studies described above, conclusions made by Sydorenko in her thesis, and by conducting pilot studies with Exploratory Factor. These hypotheses could be then tested with Confirmatory Factor analysis. Bigger samples will be required for these studies (we are currently in the process of collective big data samples from different countries; these data will also be analysed in terms of factor analysis).

As a follow-up of Sydorenko's work we will meet her supervisor, prof. Jürgen Henz at the Humbholdt University in Berlin, and organize a methodological workshop regarding the Diversity Icebreaker in the fall of this year. During this meeting we will discuss methodological issues discussed in the work of Sydorenko and the interaction between different paradigms in the questionnaire and the workshop. The ambition to gather large-data sample with the partial-ipsative format from different cultures is a consequence of Tetyana Sydorenko's work.

Content validity

The content validity of a measure refers to the degree to which this measure represents all facets of a given construct or – to be more precise – to the extent to which the items of this measure represent the empirical domain of a given construct (i.e. all possible ways of measuring this construct). A test can achieve a high content validity by careful selection of items, so that they are a representative sample of from the content domain (Anastasi & Urbina, 1997).

In order to trace the evidence related to the content validity of the Diversity Icebreaker one has to go back to the history of its creation:

The categories of Red, Blue and Green were created in the process of brain-writing (VanGundy, 1981) and by using a qualitative method of conceptualizing unstructured material (Strauss & Corbin, 1990), by focus groups (see the *1994: The creation of Red, Blue and Green* section above). Later, items were picked by hand from other, established psychological measures, by the criterion of their relevance to either of the categories; the final measure was created by discarding some items and keeping others, as to obtain the best internal consistency (see the *1997: The construction of the questionnaire* section above).

One could say that what supports the content validity of the Diversity Icebreaker questionnaire in its development process was the fact that out of 161 examples of good communication behaviours created by the focus groups, a total of 121 was then grouped and was used as a reference point for selecting items by an expert to create a measure. On the other hand, what can be interpreted as evidence not supporting the DI's content validity, as it was defined above, is the fact that 40 of these examples were left out and that the items were not grouped by factor analysis, which would establish a more definite boundaries to the categories and thus allow to better define the construct's empirical domain.

However, and most importantly, the Red, Blue and Green categories have to be considered to certain degree as emergent categories – both in relation to the Diversity Icebreaker construct validity as well as to how they are applied in the workshop.

In terms of the construct validity, the categories are emergent because there has been many explorative studies conducted providing evidence for Red, Blue and Green being related in a meaningful way to different psychological models (personality, emotional intelligence, cultural dimensions, values and most recently – cognitive styles). Furthermore, the process of globalization of the concept presents new application and research opportunities, which further expand the possible content domain of Red, Blue and Green.

In the workshop the Red, Blue and Green categories are emergent because their meaning is locally constructed and negotiated by the participants. Although they rely to greater or lesser degree on the questionnaire and often use the items as a starting point, the descriptions of the three categories are never exactly the same and bear a lot of unique elements from group to group. In other words, having a strong Blue preference, for example, may mean have a similar base but mean two different things in practice for a given workshop-group and in light of the workshop's theme.

For these reasons (the emerging character of Red, Blue and Green both in terms of the concept's construct validity and in the workshop), it may be that the notion of content validity of the

questionnaire defined as in the beginning of this section is not applicable or crucial for the Diversity Icebreaker. If the categories or construct intended to be measured in the questionnaire are not rigidly defined and no ultimate descriptions are being given, it is not possible to delimit their empirical domain and create a representative sample of items reflecting it.

This however is not a problem given the tool's application – in the workshop – where it is precisely to point not to give the ultimate descriptions of the categories, but rather invite to create a local meaning for them.

Face validity

Face validity is commonly denominated by the extent to which "the test looks like it measures, what it purports to measure". It concerns the superficial appearance, or a face value, of a measurement procedure (Gravetter & Lori-Ann, 2011). It is associated with the degree to which the test respondents view the content of a test and its items as relevant to the context in which the measure is being used (Weiner & Craighead, 2010).

Some authors claim that face validity is the simplest and least scientific aspect of a construct's validity (Gravetter & Lori-Ann, 2011). That may be so, but we see it as an important element of the Diversity Icebreaker overall validity. This is because it has a special bearing on the concept's consequential validity (which in turn supports the construct validity). It plays a crucial role in the workshop scenario, because the questionnaire is one of the sources of information for the participants creating during the group work, when they create the meaning of Red, Blue and Green.

There are two notions supporting the face validity of the Diversity Icebreaker:

First of all, the Red, Blue and Green categories emerged in what Moscovici termed the process of common sense categorization, in opposition to reified scientific methods (1984) where the random customers and not experts played the pivotal role. Thus, already in the beginning there is reason to believe that the categories are intuitive and easily identifiable, which had an effect on the questionnaire's items selection later.

Second of all, it is the questionnaire itself defines the categories, when its items are used as a starting point and inspiration for the group work in the workshop, where Red, Blue and Green categories, i.e. objects of measurement, are defined. Therefore, from the practical point of view, the measure looks like it measures what it is supposed to – because the meaning of the results of the measurement is partially defined by the questionnaire itself. As of now, we are not aware of instances where the groups failed to complete the tasks, because the questions where hard to understand or did not fit the colour they were assigned to.

However, a study of face validity of the questionnaire could be conducted to support these assumptions with a more scientific, quantitative data. Inter-rater reliability study, where independent judges rate to what degree a given item represents the category it purports to measure and the concordance of their ratings is then tested.

Consequential validity

The consequential validity is of great importance for the Diversity Icebreaker for two main reasons: firstly, the concept is intended for practical application with the specific purpose of creating a positive change persons and groups behaviours and attitudes, thus the consequences of DI's

application are of great interest to its end users and our clients; secondly, and more coherently with the paradigm and main purpose these certification assumes, "(..) appraisal of the social consequences of the testing is also seen to be subsumed as an aspect of construct validity" (Messick, 1995).

The consequential validity can be understood as one of the aspects of the overall construct validity and it is assesses the value of implications of using a measure as a basis for action as well as the actual and potential consequences of test use (Messick, 1995).

The studies below present different aspect of the consequential validity of the Diversity Icebreaker: from outcomes of a marketing campaign based on Red, Blue and Green on the macro-social level, through the concepts application in a workshop setting resulting in effects on the group level.

Some of the studies presented in this section resemble experiments (with dependent and independent variables) and could be thus considered to belong in the *External validity* section. However, the studies in question yielded results mediated by the workshop and not directly connected to the constructs of Red, Blue and Green (e.g. the effect of increased trust), i.e. it was the consequences or effects of the concept's application in the workshop that were being investigated in these experimental studies.

RED, BLUE AND GREEN IN A MARKETING CAMPAIGN FOR ENERGY REDUCTION

The categories of Red, Blue and Green emerged originally as part of the work in designing a social marketing campaign for a client and were used to a) attract attention through different media campaigns, where differentiation was made between Red, Blue and Green communication strategies; b) training advisors to reinforce seeking-contact-for-more-information behaviours in the customers; c) training advisors in giving advice for technological and behavioural change according to Red, Blue and Green preferences of the customers; and d) producing self-aid materials with arguments written from Red, Blue and Green perspectives. (See the *History of Red, Blue and Green* section above for more details.)

Ekelund's dissertation (1997) documented that the campaign reached 16 thousand customers out of 145 thousands of the county's population. The cost-benefit was evaluated and pay-back of campaign costs was measured to twenty-three million NOK in relation to estimated costs of the campaign to five million NOK. The evaluation was done through three different methods: i) comparisons between counties on energy consumption, ii) estimation of how many new, ecologically friendly, devices were installed, and iii) interviews with 96 customers, in order to estimate the effect of behavioural change and technological implementation.

Thus, there are reasons to believe that grouping preferences for communication and interaction in three categories of Red, Blue and Green, and building a set of communicational guidelines for consultants and having them trained in them, proved itself to be highly efficient and valuable for the success of the said campaign.

The results of this conservation campaign support the consequential validity of the Red, Blue and Green categories in a context where the intent is to attract attention of different segments of the population and meet them with a set of arguments, which will result in a behavioural change.

***

*(**NOTE:** The following six studies were realized within a joint research project with the Hebrew university in Jerusalem. The results and data below were presented at the POS conference at the University of Michigan, June this year. This project is part of our CSR project. You can read more about it following [this link](#).*

The present document has been written to aid the DNV validation process of the Diversity Icebreaker in Norway and for the Norwegian version of the questionnaire (questionnaires used in the studies below were either in English or Hebrew). However, in the [Convergent internal validity](#) section above correlation results between Red, Blue and Green and the Big Five personality model were reported from these studies in Israel, which were almost identical to those reported previously when the Norwegian version of the questionnaire was used in Norway. Since the Big Five personality model was reported to be stable across cultures, the abovementioned correlation results support the stability of DI model as well; thus, we state that the results of the studies below are generalizable also in Norway and hence viable for the DNV validation process.

***

EFFECTS ON AFFECT AND TRUST

In this project, the researchers made a first attempt to test the Diversity Icebreaker's potential in the area of conflict management and the workshop's immediate impact on its participants in Israel. One of the research questions was whether the DI workshop has a positive impact on affect and trust of it participants.

Expectations

The workshop was expected to produce change in both affect and cognition. Specifically, the workshop is designed to engage the participants in a fun, humorous interpersonal interaction, and create a non-judgmental environment that emphasizes the advantages of diversity. We therefore expect the workshop to increase positive affect and trust.

Study

Participants were N=211 undergraduate business students in an Israeli university (47% female; mean age = 23; 78% Israeli born, 9% Jewish immigrants, 8% Arabs). All participants were invited to take part in a workshop on interpersonal communication for partial course credit (they could choose either in English or Hebrew).

Eight 2-hour workshops were conducted within three days. The participants completed a pre-workshop questionnaire, participated in the DI procedure guided by one of three experts, and then completed a post-workshop questionnaire.

Prior to and following the DI workshop the participants completed a short questionnaire, including 20-item measures of positive and negative affect (using items from the PANAS (Watson, Clark, & Tellegen,1988), a 6-item measure of trust (adopted from Yamagishi, 1988) and a 12-item measure of willingness for contact with out-group members (adapted from Sagiv & Schwartz, 1995). Cronbach

alphas ranged from .66 to .96. All measures were split, with half measured before and half measured after the DI workshop. The order of the before and after measures was randomly reversed.

Results

The researchers compared the participants' "before" and "after" reported scores regarding affect, and distrust towards others. As expected, negative affect decreased ($t=2.42$, $p<.01$; $M_{before}=2.37$ and $M_{after}=2.29$) and positive affect increased ($t=2.56$, $p<.05$; $M_{before}=3.48$ and $M_{after}=3.57$) following the workshop. Also as expected, participants' distrust decreased following the workshop ($t=1.86$, $p<.05$; $M_{before}=4.38$ and $M_{after}=4.50$).

Personality and DI Implications

To further understand the impact of DI, the researchers investigated differential effects of the workshop on different personality types. For that aim, a set of hierarchical regressions was conducted to predict affect, distrust and tolerance following the workshop. In each regression, the "before" state was entered at the first steps, following by the DI categories (second step), personality traits (third step) and values (four step). The Stepwise method was used in all steps.

Affect. Negative affect before the workshop explained 66% of the variance in negative affect following the workshop. In addition, neuroticism predicted an increase and conscientiousness predicted a decrease in negative affect (explaining additional 3% each, $p<.05$). Positive affect prior to the workshop explained 25% of the variance in positive affect following. Agreeableness explained 4% more (predicting increase in positive affect). Surprisingly, benevolence values predicted a decrease in positive affect (explaining 6%). This effect disappears when only values are regressed on the change in positive affect.

Distrust. Distrust prior to the workshop explained 36% of the variance in distrust following. In addition, power values and neuroticism predicted increased in distrust (i.e., low power and emotional stability predicted decrease in distrust, explaining 5%, $p<.01$ and 2%, $p<.05$ respectively).

Discussion

The effects of the Diversity Icebreaker workshop described above have a profound meaning for the concept's consequential validity. They confirm in a controlled, quasi-experimental setting, what we have been seeing in the workshops and what our clients and the workshop participants have been reporting. The positive and collective affective experience that workshop conjures up is one of the unique qualities of this tool.

Furthermore, the increase in positive affect and trust, as well as the decrease in negative effect achieved within such a short time span, in a predictable and standardized workshop-scenario, provides a strong argument for applying the workshop in conflict-settings. In such settings, it is crucial to reduce mistrust and negative affect as soon as possible, in order to be able to go into a meaningful dialogue, and the DI workshop can be a way of doing that.

EFFECTS ON CREATIVITY AND SOCIAL IDENTITY

Within the same research project, there was one more the general question raised of whether the practices that have positive influence on employee's performance and business goals achievement

can be flexible and whether they can be shaped in an organizational intervention. The researches focused on interpersonal interactions and creativity as examples of areas of these practices and proposed the Diversity Icebreaker as an intervention that can have a positive effect on both.

Expectations

The researches hypothesised that the Diversity Icebreaker workshop will facilitate creativity and awareness for social identity among its participants.

Method

The sample was N=82 participants (46% female, mean age = 24.), who were assigned into four workshops. Creativity was assessed using a divergent thinking task (Wallach & Kogan, 1965). The Twenty Statement Test (Kuhn & McPartland, 1954) assessed the participants' identity. The order of the creativity task versus the identity task was manipulated: Participants were randomly assigned into "creativity-first" versus "identity-first" condition.

Results

As expected, the intervention facilitated originality. The ideas provided by the participants after the workshop were more rare than those provided prior the workshop $t(79)=2.11$, $p<.05$. Also as hypothesized, participants described themselves more in relational terms after the intervention than prior to the intervention $t(79)=1.95$, $p<.05$. Importantly, the extent to which participants described themselves in individualistic terms did not change following the workshop.

In sum, this study provides support for the Diversity Icebreaker's consequential validity in terms of increasing creativity and the relational-identity (relational-Self) of the workshop's participants; in other words: the workshop simultaneously facilitated both autonomy (leading to personal creativity) and social engagement.

However, the researchers note that future studies could focus on long-terms outcomes of the workshop in real-life organizational settings.

SATISFACTION WITH THE DI WORKSHOP
In addition to the variables measured in the study about the impact of DI on affect and trust (described above) one of the pre-/post-measures was aimed at measuring the participant's satisfaction with the Diversity Icebreaker workshop. This study is presented in this section, since the overall satisfaction with the workshop is of importance for the Diversity Icebreaker application. We assume that the positive experience will function as a reinforcer/reward (as in the classical operant conditioning), increasing the probability of the positive behaviour in the future, e.g. regarding being open and self-reflective when engaged in dialogue processes.

Results

The researchers report that overall, all eight workshops were successful in terms of active participation and quality of interaction. In addition to gathering this non-formal information from the participants, the researchers asked them to rate their satisfaction with the workshop on a 7-point scale. The findings reveal that the workshop has been interesting (6.43) and enjoyable (6.25), and to a somewhat less extent fulfilling (5.04) and useful (5.33). In contrast, is has not been viewed as

Documentation for description and evaluation of the Diversity Icebreaker

difficult (1.83) nor irritating (1.43). Thus, it was concluded that the participants were highly satisfied with the workshop.

DIFFERENCE BETWEEN FIRST TIME USERS AND EXPERIENCES DI USERS

In the study about DI's impact on affect and trust (N=211), within the same research project realized with the researchers in Israel, also the difference in terms of change in the positive affect and distrust was measured in function of the person of the facilitator.

Overall, there were three facilitators: one experienced, English-speaking – Bjørn Z. Ekelund, and two first-time Hebrew-speaking facilitators, who had read the User Material and got acquainted with other instructional material for consultants.

The results are presented in Table 23 through 26, and Figure 10 and 11 on the next pages.

Table 27. *Positive affect before and after the workshop for three facilitators - descriptives*

	Facilitator	Mean	Std. Deviation	N
PA_before	Bjorn	3.4800	.50429	105
	Rotem	3.5040	.49213	59
	Tammy	3.4511	.48895	47
	Total	3.4803	.49553	211
PA_after	Bjorn	3.5319	.47267	105
	Rotem	3.6489	.60571	59
	Tammy	3.5713	.43498	47
	Total	3.5734	.50585	211

Table 28. *Tests of Within-Subjects Contrasts - PA*

Source	factor1	Type III Sum of Squares	df	Mean Square	F	Sig.
PA	Linear	1.052	1	1.052	7.510	.007
PA* Facilitator	Linear	.186	2	.093	.662	.517
Error(PA)	Linear	29.151	208	.140		

Figure 10. *Estimated Marginal means for PA (factor 1) before (blue line) and after (green line) for the three facilitators (Bjørn, Tammy and Rotem)*
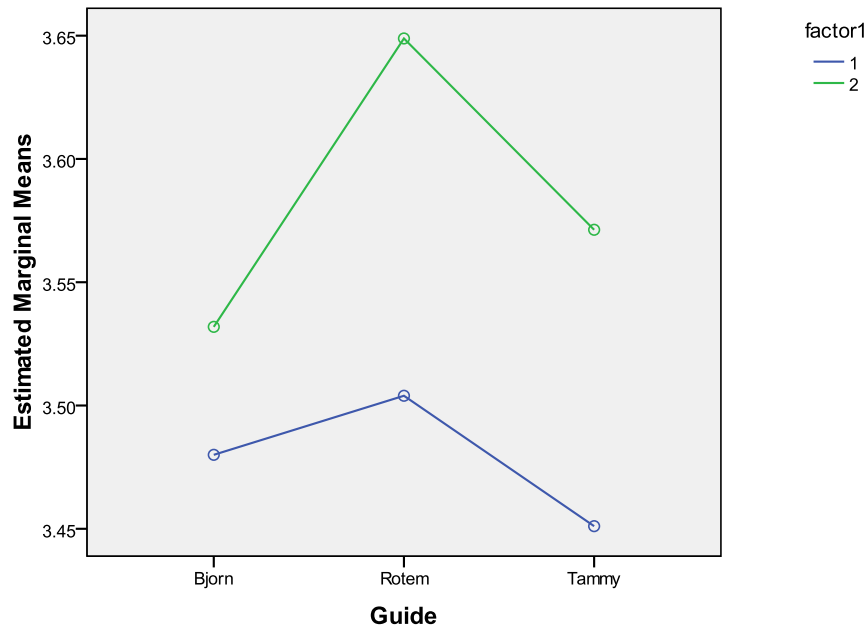
Table 29. *Distrust before and after the workshop for three facilitators - descriptives*

	Guide	Mean	Std. Deviation	N
distrust_before	Bjorn	3.6540	1.31877	105
	Rotem	3.5367	1.16628	59
	Tammy	3.7447	1.14675	47
	Total	3.6414	1.23716	211
distrust_after	Bjorn	3.5556	1.20599	105
	Rotem	3.4576	1.13276	59
	Tammy	3.4681	1.16441	47
	Total	3.5087	1.17212	211

Table 30. *Tests of Within-Subjects Contrasts - distrust*

Source	Distrust	Type III Sum of Squares	df	Mean Square	F	Sig.
Distrust	Linear	2.159	1	2.159	4.012	.046
Distrust * facilitator	Linear	.633	2	.317	.588	.556
Error(Distrust)	Linear	111.954	208	.538		

Figure 11. *Estimated Marginal means for Distrust (factor 1) before (blue line) and after (green line) for the three facilitators (Bjørn, Tammy and Rotem)*
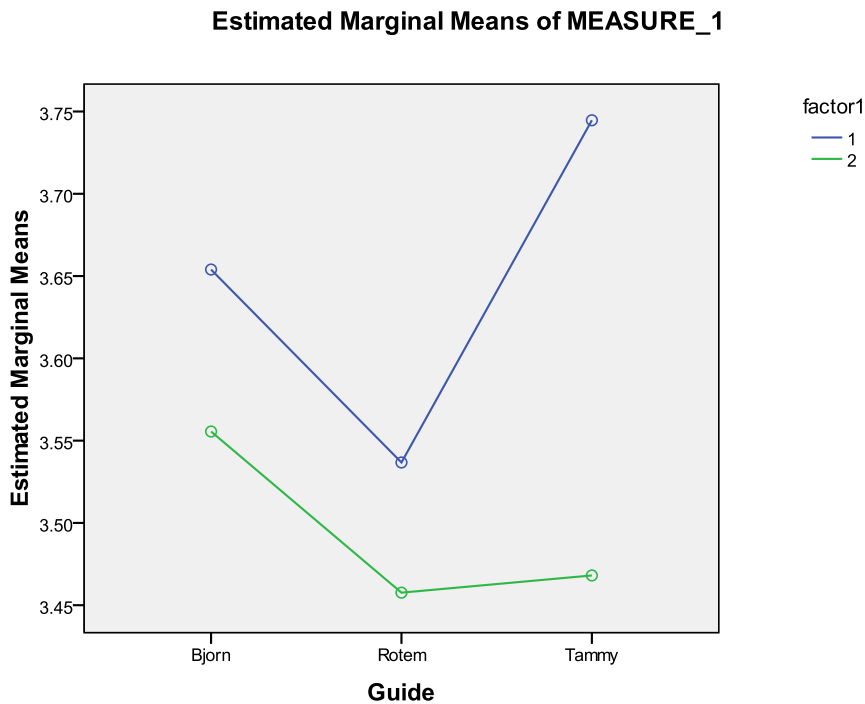
The outcome evaluation of the workshop is better for first time instructor in Hebrew (Tammy and Rotem) than experienced English speaking consultant (Bjørn). The possible explanation is that the participants could have responded better to the workshops in Hebrew (being their primary language) than in English.

However, the differences are not significant (p=.517 for PA and p=.556 for Distrust).

The results indicate that our training material for consultants (the User Manual, DVD, etc.) is functioning to a satisfactory level in preparing first time user to conduct the workshop.

DI HAS NO EFFECTS ON THE GENERAL SENSE OF TOLERANCE

This study is also a part of the same research project (described above), with the Hebrew University in Jerusalem. It can be viewed as supporting the "divergent consequential validity" of the concept, i.e. indicating areas where the DI workshop has no effects.

In this study, the researchers from the Hebrew University in Jerusalem together with Bjørn Z. Ekelund investigated the Diversity Icebreaker's impact on tolerance, specifically – the authors wanted to test whether it could increase the willingness to work with the Ultra-Orthodox Jews. It was hypothesized that if the DI workshop increases positivity to diversity it would also increase willingness to work with Ultra-Orthodox Jews (a group adding to the diversity in the country of Israel).

Documentation for description and evaluation of the Diversity Icebreaker

<u>Study</u>

Participants were N=211 undergraduate business students in an Israeli university (47% female; mean age = 23; 78% Israeli born, 9% Jewish immigrants, 8% Arabs). All participants were invited to take part in a workshop on interpersonal communication for partial course credit (they could choose either in English or Hebrew), i.e. same group as previously.

Eight 2-hour workshops were conducted within three days. The participants completed a pre-workshop questionnaire, participated in the DI procedure guided by one of three experts, and then completed a post-workshop questionnaire.

Prior to and following the DI workshop the participants completed a short questionnaire measuring their willingness for contact with out-group members (adapted from Sagiv & Schwartz, 1995). Measures were split, with half measured before and half measured after the DI workshop. The order of the before and after measures was randomly reversed.

<u>Results</u>

The 12-item scale measuring willingness for contact with out-group had an internal reliability of Cronbach's α=.92.

A dependant sample t-test was conducted to compare the participants' tolerance towards out-group before and after the Diversity Icebreaker workshop. There was a significant difference between the scores before M=5.00 and after the workshop M=4.90, t(210)=3.06, p=.01. These results show that the willingness to have contact with the Ultra-Orthodox Jews group has decreased after the workshop, contrary to the expectations.

<u>Discussion</u>

Following is the possible, post-experimental explanation of the obtained results: in the Diversity Icebreaker an image of good practice for interaction across the diversity is created among and by the workshop participants. Since none Ultra-Orthodox Jews took part in the workshop, one might expect that they are not privy to this attractive image of good practice for interaction and thus even less attractive for the participants, being an out-group.

Furthermore, the Ultra-Orthodox Jews may have been perceived by the participants as a generally not very open or inclusive group by themselves. Thus if the participants felt that their personal tolerance has increased after the workshop, it may be that they perceived the Ultra-Orthodox Jews group as even less similar to themselves and unwilling to relate to them. In the future, it could be interesting to measure the willingness to work with an out-group by choosing another, less politically controversial example.

This study bears certain consequences for the external divergent validity of the Diversity Icebreaker. It indicates that it is not a tool for a general increase of tolerance and trust towards any given out-group – although it increases trust on the in-group level, i.e. between the participants taking part in the workshop (see the *Consequential validity* section for studies supporting that claim).

Moreover, these observations have consequences for the tool application, which we have already partially addressed in our practice: a) in the debriefing stage of the workshop, we explicitly name the

threats related to labelling and stigmatization; b) we encourage the participants to be inclusive towards members of their organization or unit, who they work with and who could not take part in the workshop, share the learning points from the workshop with them and provide them with additional copies of the questionnaire; and c) show the value to our clients of including the Diversity Icebreaker in trainings for whole units, departments and organizations.

EFFECTS ON COMMUNICATION IN HETEROGENEOUS AND HOMOGENOUS TEAMS
A study described in a master thesis by Nordgård has evaluated two training sessions delivered to a heterogeneous and a homogenous team (in terms of different professional backgrounds or lack of thereof). Improved communication in the homogenous team was documented (Nordgård, 2008).

The main purpose of the study was to investigate whether application of the Diversity Icebreaker could improve communication in these two teams. Qasi-experimental design was chosen to test the assumptions of a positive effect of the Diversity Icebreaker on communication in teams. The study was quasi-experimental in the sense that there was a pre-test and post-test of the variables, but there was no control group and no randomization in the sample selection (the quality of communication was controlled prior to the Diversity Icebreaker application and afterwards).

Interviews and questionnaire methods were used to measure the variables (quality of communication) and the data was analysed using the Ground-theory approach.

The homogenous team was characterized by a worse communication than the heterogeneous one before the intervention. After the Diversity Icebreaker workshop there was observed an improvement in communication in the homogeneous team. The study also showed that the feeling of security was an important factor for the quality of communication in the teams.

DI AND FLOW
Lisa Vivoll Straume (Straume & Ekelund, 2005) measured the degree of FLOW defined by Csikszentmihaly (2003) using a simplex model (feeling of happiness, easiness, joy, challenging, dramatic) developed by Vittersø, (1999). This was done following six classic DI workshop (total of N=127 participants).

All participants independent of Red, Blue and Green scored extreme high on the FLOW dimensions, independently of the dominant preference. This result confirmed what could have been expected based on the feedback received from participants and consultants using DI.

It would appear that the workshop one can create a certain atmosphere promoting FLOW in a standardized (in line with the scenario) way. We have experienced that following such exercise people become open and easily share own perceptions and prejudices, of one-self and of the other, with humour being a pivotal element. In such way we create a shared, collective experience that creates trust linked to diversity and communication.

Canney-Davison & Ekelund (2004) write that the collective identity is built by having a positive peak-experiences shared collectively by the group; and that this experience promotes the emergent states of trust. In our view, the workshop creates such "peak-moments" and provides the participants with a collective feeling of being able to succeed individually and as a group. The ability to address diversity in a constructive way is an integrated part in these collective experiences is.

EFFECTS ON FLEXIBILITY IN ROLE TAKING

Among the qualities of the Diversity Icebreaker model and workshop that can increase flexibility in behaviour and interaction, is the fact that the Red, Blue and Green are not independent (see the *Factor structure* section above), which is reflected in the partial-ipsative format of the questionnaire. This allows the participants to describe themselves in line with e.g. "I am primarily Red, but also Blue and Green to a certain extent". During the workshop they further realize that they can use this notion and their different preferences, even the less dominant on most of occasions, in order change their behaviour in function of situation or to better relate to a partner in an interaction.

This flexibility is often tested already during the classic Diversity Icebreaker workshop, at the time when the facilitator splits the group into even numbered groups of Red, Blue and Green. Most of the workshop groups do not represent the norms ideally and are skewed (e.g. the IT workers tend to score high on Blue), which implies that some of the participants will be asked to work in a colour-group that is not the same as their most dominant colour. We then explain however, that in the given workshop setting, at that time, this person is still among "the most Red participants" of all the participants, for example (which is assured by the process of dividing the participants), despite being "primarily Blue". We further encourage to take on the "Red-role" by saying that he or she have obtained some scores on all of the three colours, and he or she will be able to use the non-dominant preference in most situations.

A study by Andrey Elster and Lilach Sagiv (the Hebrew University of Jerusalem, Israel) in the spring of 2012 tested this assumption by comparing if there was a difference between the workshop-groups where the participants were assigned to the Red, Blue and Green work-groups in a typical way and workshop-groups where they were assigned randomly.

Study

In their study, N=90 business school students (M_{age}=23, 46% females) participated in 5 different DI workshops. The positive (PA) and negative affect (NA) were measured before and after the workshop (the *PANAS* measure was used; as in the study described under *Effects on affect and trust* above, in the present section). Half of the participants (N=40) were randomly divided into colour-groups while the other half (N=50) were divided based upon their questionnaire results.

The paired sample t-test showed same results as the former studies within this project, i.e. an increase in the PA and Trust, and decrease of NA for both groups.

Table 31. *Means for NA, PA and Trust before and after the workshop for both groups (N=90)*

	Before	After	t	df	p value
Positive affect	3.37	3.54	2.78	89	.007
Negative affect	2.35	2.19	2.85	89	.005
Trust	4.24	4.49	2.42	89	.017

Furthermore, the PA, NA and Trust variables changed in the same way and directions in both random and non-random groups, which means the assignment according to the dominant preference or lack

of thereof had no effect on the overall workshop effect. Below are the graphical illustrations of these results.

Figure 12. *Change in negative affect for random and non-random workshop groups, before and after the workshop (N=90).*
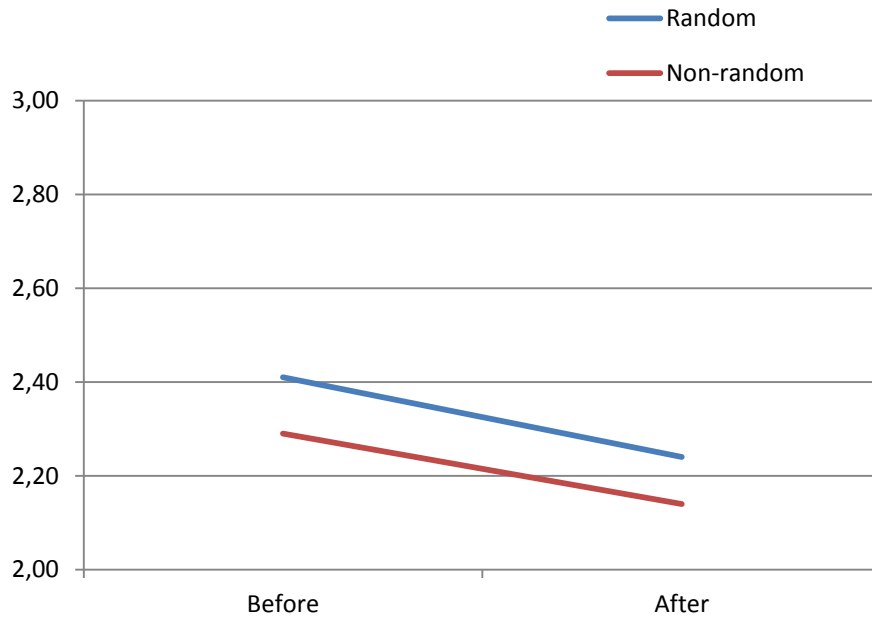
Figure 13. *Change in positive affect for random and non-random workshop groups, before and after the workshop (N=90).*

Documentation for description and evaluation of the Diversity Icebreaker
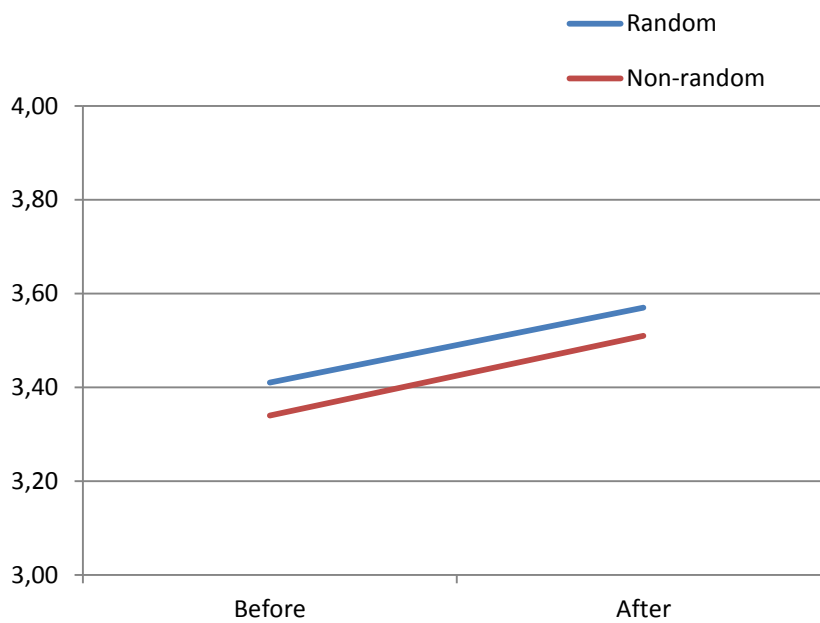
Figure 14. *Change in trust for random and non-random workshop groups, before and after the workshop (N=90).*
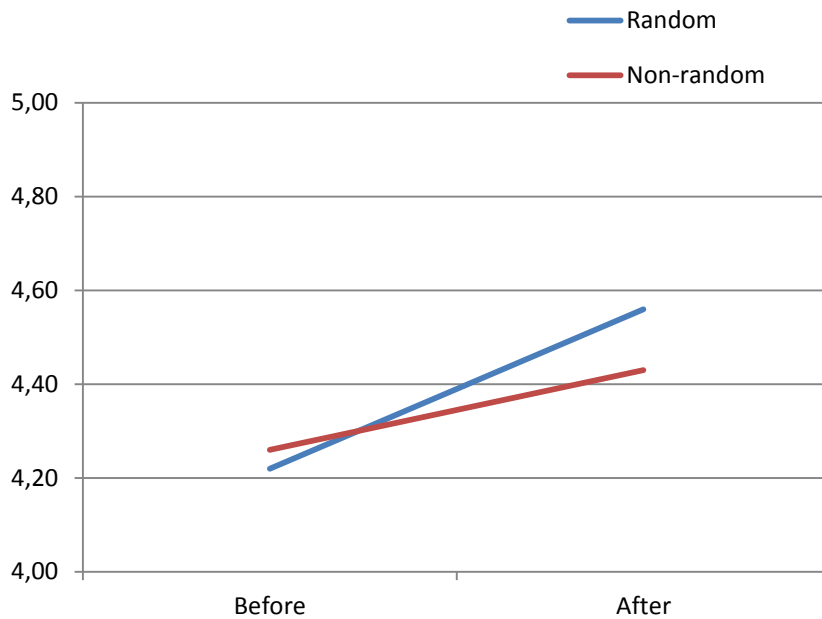
Independent samples t-test – the feeling of fit and workshop evaluation

However, there was one variable measured that differed the groups significantly: The participants were asked whether they felt the categories fit for them, and the non-random group had higher degree of feeling of fit.

Table 32. *Means for the feeling of group fit for random and non-random group (N=90)*

	Random	Non-random	t	df	p value
Group-fit	4.80	5.67	2.62	89	.011

Furthermore, also the ratings of the workshop by the participants in terms of its usefulness, interestingness and enjoyableness were compared between the groups. The participants that were in the non-randomly assigned groups rated the workshop higher on all of these elements:

		Group fit
Workshop	**Interesting**	.23*
	Enjoyable	.23*
	Useful	.18[a]

USER EXPERIENCES (QUALITATIVE ANALYSIS)

Qualitative analyses, with participation of 4 very experience consultants that have been using the concept extensively, have been conducted (Ekelund & Langvik, 2008). Results were obtained with the use of grounded analysis and a software package named Nvivo 7.0. The results have been presented at Academy of Management Annual conference in 2007 and in the book from 2008 (Ekelund & Langvik (2008). The results of the study highlighted six areas, which highly integrated in practice:

  a) A user friendly instrument, intuitive categories

  b) Creates an emotional event characterized by positive affect

  c) Offers a new language and shared understanding to manage diversity

  d) Dynamic polarization

  e) Creates self-, other- and team-knowledge

  f) Facilitates cooperation in organisations

CUSTOMER SURVEY

In addition to these results a larger-scale, qualitative data has been gathered in a market-survey among the DI users in the autumn of 2012.In this place at the end of this section we would like to present its results. It is to illustrate how the users of our tool evaluate it and where and how they apply it, and thus provide additional, end-process evidence regarding the consequential validity of DI from the real application-context.

The survey was sent to N=373 Norwegian customers on our newsletter list - ; a total of N=127 answered the survey (34% response rate). The general purpose of the survey was to learn about the possible improvements and future development areas. Furthermore, we wanted to learn in which of areas of the tool's application are the most popular ones and hence which of them should we focus on most (in terms of developing customer-support and research).

The majority of the respondents have used DI more than once: 40% 2 to 5 times and 46% over 6 times; 91% of them were either extremely satisfied (*svært godt fornøyd,* 39%) or very satisfied (*godt fornøyd*, 52%):

Figure 15. *Totalt sett, hvor fornøyd eller misfornøyd er du med Diversity Icebreaker som produkt?*
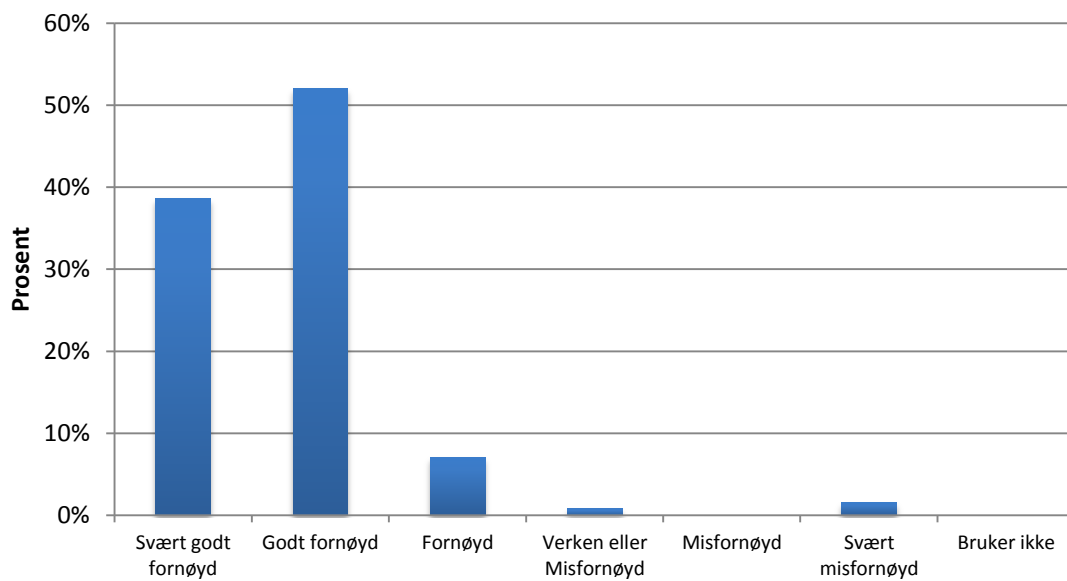
This implies that there are many recurring DI customers and that a great part of them is very satisfied with the tool – providing "real-market" evidence supporting the effects of the tool descried in research studies in this section above.

Furthermore, we asked the respondents to indicate within which areas of application they used the concept most frequently. The results are plotted in the graph below:

Figure 16. *Within which areas do you apply the Diversity Icebreaker? (You can mark more than one area).*
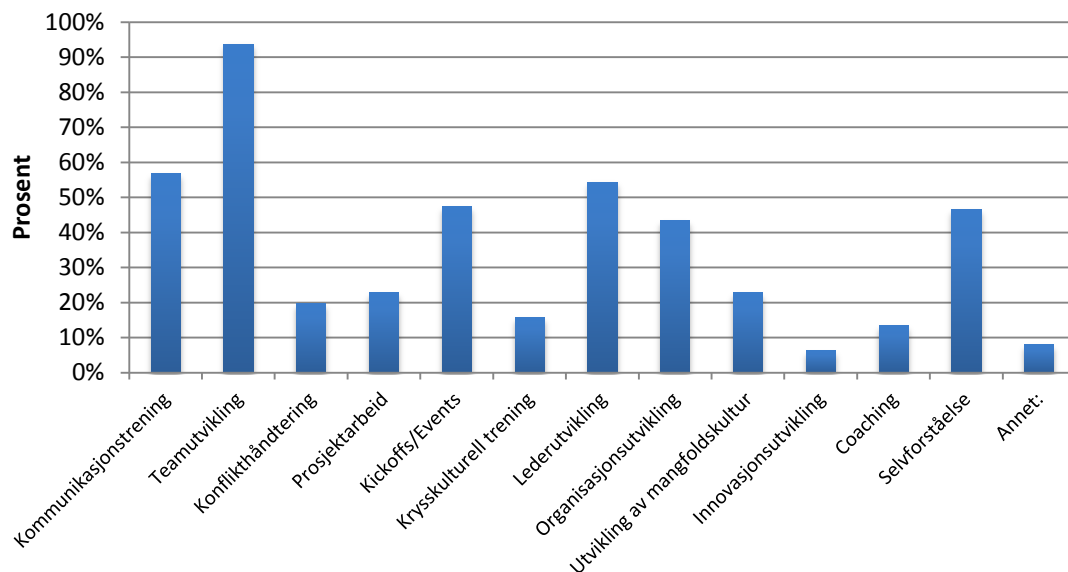
These results confirm that the tool is used primarily in the area of team-building, which confirming its classification as a team-tool. Historically, the DI categories and interaction between people with different colour preferences, draws on ideas from team role concepts like the Belbin and Margerison & McCann (TMP) models. It is also so that it was presented first time as an alternative teamrole model in 1998 in certification training for Team Climate Inventory (Ekelund & Jørstad, 2002). For this reason it is no wonder that this has become the most prevalent user area for DI.

On-going research projects

CREATING SHARED LANGUAGE IN AN ORGANIZATION

Work in progress: In order to broaden the research on experiences following use of Diversity Icebreaker in organization we decided to start a qualitative research design including a within-methods triangulation, 4 different types of organizations were chosen in Norway that have used the concept in different ways. These include a construction company that has used the DI concept for 5 years for project development, leadership programs and kick-offs of large projects; an industrial group that has used it in leadership and team development; a hospital that has used it for increasing attractiveness for teamwork; and a local municipality that has used it for intercultural interaction, leadership development, interpersonal interaction and as a platform for embracing diversity in general ways. The data for this project consists of eight in-depth interviews and two extended focus group analysis sessions. Data were transcribed and analysed by the two interviewers separately (one, a native of Norway, the other not) and further cross-checked by two psychology students (one, Norwegian, one not) in separate processes. A preliminary model of a categorical system agreed upon among the involved researchers based upon the main statements from the interviews look like the model below.

Figure 17. *Model of the DI workshop as seen by users*, (Ekelund, S.M. & Ekelund, B.Z., 2012)
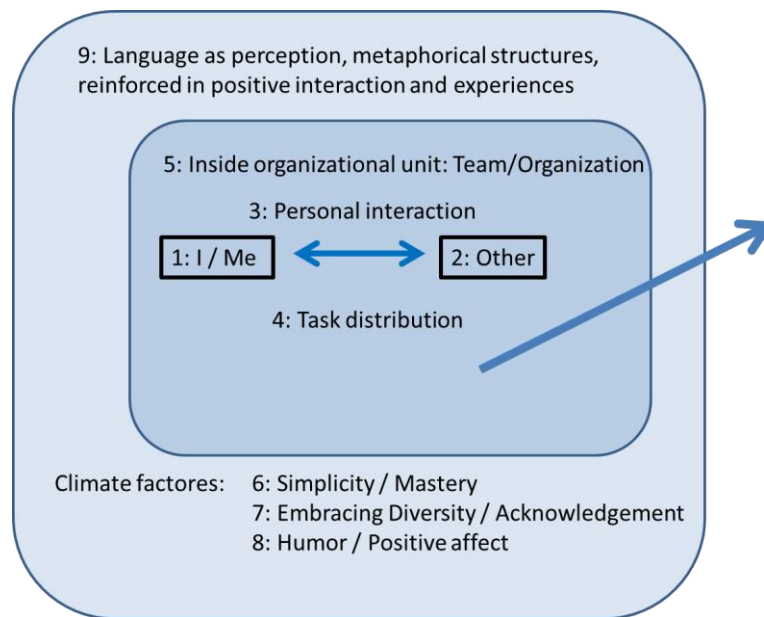
THE DIVERSITY ICEBREAKER HUMOUR PROJECT

We have long since recognized the important role humour plays in the Diversity Icebreaker seminars. However, it is the first time that we took on a systematic and a scientific approach to investigate how humour contributes to making people *reflect, ask, listen and talk* during our seminars.

Following are the reasons for which we want to investigate humour in DI:

The reported, considerable amount of humour in DI should be reason enough to believe that it plays an important, if not crucial, role in the seminars. In other words, since participants engage in laughter so often and so eagerly, it must be to a purpose and it must have some serious social functions.

Second of all, it's *repeatable*, meaning that the DI provides a consultant-proof scenario to follow, which always seems to create laughter independently of particular group's characteristics and/or those of the consultant.

Third, a consultant may initiate and participate in the shared humour experience during the seminar, but it is the participants who generate *most* of humour and laughter. Therefore we can take a look at humour-in-making in real social interactions – a setting most appropriate to study humour but rarely approachable.

In general, we want to focus on two areas of investigation in this project: a) what are the different humour styles (Martin, Puhlik-Doris, Larsen, Gray, & Weir, 2003) related to either Red, Blue or Green and b) what are the effects of humour and through which mechanism social and cognitive mechanism they come into play.

In regard to latter area of investigation, we have hypothetically delimited the possible functions humour in the seminars, drawing on an extensive body of theory and research from within the field of psychology of humour (Martin R. A., 2007): *tension relief, increasing liking, cohesion and identity building, enhancing trust, breaking norms* and *enhancing self-understanding.*

(For more information on this project refer to the *Diversity Icebreaker humour project - Piotr Pluta* file.)

USE OF LEARNING STYLES IN MULTICULTURAL STUDENT GROUPS

Together with Marieke von Egmond and Alexis Rossi we work on gathering data about differences in learning styles between East and West, combined with DI dimensions. The project focuses on analysis of Pedagogical examples and illustrations from classroom culture (pedagogical). The first article, a result of this work, has been presented on the Democracy and Diversity in Education Conference at Høgskolen i Buskerud earlier this year (Rossi, et al., 2013).

DI IN BUILDING OF THE THIRD CULTURE

Bjørn Z. Ekelund together with Kazuma Matoba write a chapter for a forthcoming book titled "Beyond Hofstede". The chapter will integrate the Diversity Icebreaker concept of egalitarian categories and shared language created under the seminar, into Kazuma's model of cosmopolitan communication – the Third Culture.

CROSS-CULTURAL VALIDATION

Rotem Shneor and Human Factors AS are gathering data from different countries to conduct a cross-cultural validation of the DI questionnaire. We expect to obtain 400-500 respondents from the following countries: Israel, Italy, Germany, Turkey and Norway. We also intend to use these data to discuss and explore further the properties of the partial ipsative response format in the DI questionnaire.

RED, BLUE AND GREEN AND THINKING STYLES

After the preliminary study where the Post's measure of thinking styles (2011) have been employed (see the *Red, Blue and Green and two thinking styles* section above), we have begun to gather data using using a more advanced measure of cognitive styles – the Thinking Styles Inventory (Zhang & Sternberg, 2006). This project is aimed to advance our understanding of the DI model as a potential cognitive diversity model.

List of tables

List of figures

Bibliography

Anastasi, A., & Urbina, S. (1997). *Psychological testing.* Upper Saddle River, NJ: Prentice-Hall.

Aronson, E., Wilson T, D., Akert, R. M., & Fehr, B. (2007). *Social psychology.* Toronto: Pearson Education.

Bartman, D., Lindley, P., & Kennedy, N. (2004, November). EFPA Review Model for the Description and Evaluation of Psychological Tests. Test Reviiew Form and Notes for Reviewers.

Belbin, M. (1981). *Management Teams: Why they suceed or fail?* Oxford: Heinemann.

Benson, J. (1998). Developing a Strong Program of Construct Validation: A Test Anxiety Example. *Educational Measurement: Issues and Practice*, 10-22.

Benson, J., & Hagtvet, K. A. (1996). The interplay among design, data analysis, and theory in the measurement of coping. In M. Zeidner, & N. S. Endler, *Handbook of Coping.* New York: John Wiley & Sons, Inc.

Bjørn, Z. E. (1997). The application of a model which integrates market segmentation and psychological theories to change energy consumption in households. Dissertation MBA. Henley Management College/Brunel University.

Block, F. (2012). *Observational validation of the Diversity Icebreaker - a first step.* Studen work submitted to the Friedrich-Schiller University of Jena, Department of Intercultural Business Communication.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*, 81-105.

Canney-Davison, S., & Ekelund, B. Z. (2004). Team processes in global teams. In H. Lane, M. Mendenhall, M. L. Maznevski, & J. McNett (Eds.), *Handbook of Global Management. A Guide to Managing Complexity.* (pp. 227-249). Hoboken, New Jersey: Blackwell Publishing.

Costa, P. T., & McCrae, R. R. (1988). Personality in Adulthood: A Six-Year Longitudinal Study of Self-Reports and Spouse Ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology, 54*(5), 853-863.

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R: Professional manual.* Odessa, Florida: Psychological Assessment Resources.

Costello, A. B., & Osborne, J. W. (2005, July). Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Crocker, L. M., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Holt, Rinehart, and Winston.

Cronbach, A. B., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*(52), 281-302.

Csikszentmihalyi, M. (2003). *Good Business. Leadership, Flow and Making of Meaning.* London: Hodder and Stoughton.

Culler, J. D. (1976). *Ferdinand De Saussure.* Glasgow: Willaim Collins Sons & Co.

Dunlap, W. O., & Cornwell, J. M. (1994). Factors analysis of ipsative measures. *Multivariate Behavioural Research, 29*(1), 115-126.

Ekelund, B. Z. (1983). Validering av et livskvalitetsmål for bruk i spørreskjema. Hovedoppgave. Psykologisk Institutt, Universitetet i Oslo.

Ekelund, B. Z. (2009). Håndtering av forskjellighet i team (Managing diversity in teams). In H. Fyhn, *Kreativ tverrfaglighet, teori og praksis* (pp. 182-205). Oslo: Tapir forlag.

Ekelund, B. Z. (2010, August 30). User Manual. Facilitating Diversity Icebreaker seminars. Oslo, Norway: Human Factors AS.

Ekelund, B. Z., & Jørstad, K. (1998). Team Climate Inventory - an intervention manual. Oslo, Norway: Human Factors AS.

Ekelund, B. Z., & Jørstad, K. (2002). *Team Climate Inventory intervention manual (Danish).* Copenhagen: Danish Psychological Publisher.

Ekelund, B. Z., & Langvik, E. (2008). *Diversity Icebreaker: How to Manage Diversity Processes.* Oslo: Human Factors Publishing.

Ekelund, B. Z., & Rydningen, M. (2008). Diversity Icebreaker - Personal Workbook. Oslo: Human Factors Publishing.

Ekelund, B. Z., Davcheva, L., & Iversen, J. V. (2009, December). Diversity Icebreaker: Developing share understanding of cooperation (accepted for publication). *Bulgarian Journal of Psychology*.

Ekelund, B. Z., Rossi, A., & van Egmond, M. C. (2010). Use of Diversity Icebreaker and learning styles in multicultural teaching settings (workshop). *The European Conference on Educational Research.* University of Helsinki, Finland.

Ekvall, G., & Arvonen, J. (1991). Change-centered leadership: An extension of the two-dimensional model. *Scandinavian Journal of Management, 7*, 17-26.

Furham, A., & Petrides, K. V. (2003). Trait Emotional Intelligence and Happiness. *Social Behavior and Personality, 31*(8), 815-824.

Gardner, H., & Martinko, M. J. (1996). Using the Myers-Briggs Type Indicator to study managers: A literature review and research agenda. *Journal of Management, 22*, 45-83.

Gravetter, F. J., & Lori-Ann, B. F. (2011). *Research Methods for the Behavioral Sciences.* Cengage Learning.

Hall, E. T. (1959). *The Silent Language.* New York: Anchor Press.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel psychology, 41*, 43-62.

Hegge, T. I. (1997). *Rapport om arbeidet med segmenter, utprøving av kommunikasjonsbiter rød, blå og grønn og utprøving av kWh-kur.* Rapport til Akershus Eneriverk.

Hjertnes-Schjødt, B. R. (2013, Mai). Samfunnspsykolog, allmennpsykolog eller kommunepsykolog - eller bare en helt alminnelig psykolog? *Tidsskrift for Norsk Psykologforening, 50*, 426-433.

Hofstede, F. (2001). *Culture's consequences. 2nd edition.* Thousand Oaks: Sage Publ.

Honey, P., & Mumford, A. (1992). *The manual of learning styles.* Berkshire: Maidenhead.

House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, Leadership and Organizations: The GLOBE Study of 62 Societies.* Thousand Oaks, CA: Sage.

Innvær, S. (1999). *"Grid-Group"-analysen i lys av weberiansk metodologi. (Dissertation dr.polit.).* Bergen: University of Bergen.

Jabri, M. M. (1991). The Development of Conceptually Independent Subscales in the Measurement of Modes of Problem Solving. *Educational and Psychological Measurement, 51*(4), 975-983.

Jones, E. E., & Nisbett, R. E. (1972). The Actors and Observer: Divergent Perspectives of the Causes of Behavior. *Journal of Personality and Social Psychology, 27*(2), 79-94.

Kirkman, B. L., Lowe, K. B., & Gibson, C. B. (2006). A Quarter Century of Culture's Consequences: A Review of Empirical Research Incorporating Hofstede's Cultural Values Framework. *Journal of International Business Studies, 37*(3), 285-320.

Kluckhohn, F. R., & Strodtbeck, F. L. (1961). *Variations in value orientations.* Wesport, CT: Greenwood.

Kuhn, M. H., & McPartland, T. S. (1954). An empirical investigation of self-attitudes. *American Sociological Review, 19*, 68-76.

Langvik, E. (2006, July 22-26). Personality Traits and Team Roles: Introducing a Tricolour Model of Team Roles and its Relationship to Personality Traits in the Five Factor Model. *The 13th European Conference on Personality*. Athens.

Langvik, E. (2009, July 10). The ipsative format of Diversity Icebreaker. *Presentation at the 11th European Psychological Conference*. Oslo.

Langvik, E., & Ekelund, B. Z. (2006, November 9). Managing Diversity: How to heterogenize teams in order to create self- and team-knowledge. *The 8th conference on Social and Community Psychology*. Trondheim: NTNU.

Margerison, C., & McCann, D. (1991). *Team Management. Practical Approaches.* London: Mercury Books.

Martin, A. R., Puhlik-Doris, P., Larsen, G., Gray, J., & Weir, K. (2003). Individual differences in uses of humor and their relation to psychological well-being: Development of the Humor Styles Questionnaire. *Journal of Research in Personality*(37), 48-75.

Martin, R. A. (2007). *The psychology of humor: an integrative approach.* Burlington: Elsevier Academic Press.

Matoba, K. (2011). *Transformative Dialogue for Third Culture Building. Integrated Constructionist Approach for Managing Diversity.* Opladen: Budrich UniPress.

Maznevski, M. L. (194). Understanding Our Differences: Performance in Decision-Making Groups with Diverse Members. *Human Relations, 47*(5), 531-552.

McCann, D. (1988). *How to influence others at work.* London: Heinemann.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*(9), 741-749.

Moscovici, S. (1984). The phenomenon of social representations. In R. M. Farr, & S. Moscovici, *Social representations* (pp. 3-69). Cambridge: Cambridge University Press.

Murphy, K. R., & Davidshofer, C. O. (2002). *Psychological Testing. Principles and Applications.* (5th ed.). New Jersey: Prentice Hall.

Nysæter, T. E., Langvik , E., Berthelsen, M., & Nordvik, H. (2009). Interpersonal problems and personality traits. The relation between IIP-64C and NEO-FFI. *Nordic Psychology, 61*(3), 82-93.

Petrides, K. V., & Furnham, A. (2001). Trait emotional intelligence: Psychometric investigation with reference to established trait taxonomies. *European Journal of Personality, 15*(6), 425-448.

Petrides, K. V., & Furnham, A. (2001). Trait Emotional Intelligence: Psychometric Investigation with Reference to Established Trait Taxonomies. *European Journal of Personality, 15*(6), 425-448.

Post, C. (2011). Relating Cognitive Style Composition, Mobility and Faultlines to Team Innovation. *Academy of Management Annual Meeting (submission accepted)*, (pp. 1-37).

Rommetvedt, R. (1972). *Srpåk, tanke og kommunikasjon: ei innføring i språkpsykologi og psykolingvistikk.* Oslo: Universitetsforlaget.

Rossi, A., Van Egmond, M., Orgeret, K. S., Pluta, P., & Ekelund, B. Z. (2013). Promoting democratic practice through 'Diversity Icebreaker' in multicultural student. *Democracy and Diversity in Education (CD).* Drammen: Buskerud University College.

Schwartz, S. H. (1992). Universals in the Content and Structure of Values: Theory and Empirical Tests in 20 Countries. In M. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 25, pp. 1-65). New York: Academic Press.

Schwartz, S. H., & Sagiv, L. (1995). Identifying culture-specifics in the content and structure of values. *Journal of Cross-Cultural Psychology, 26*, 92-116.

Seligman, M. E., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55*(1), 5-14.

Soldz, S., & Vaillant, G. E. (1999). The Big Five Personality Traits and the Life Course: A 45-year Longitudinal Study. *Journal of Research in Personality*(33), 208-232.

Straume, L. V., & Ekelund, B. Z. (2005, September). FLOW following introduction of a diversity tool. *Positive Psychology Conference, NPF*. Oslo, Norway.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research.* London: Sage Publications.

Sydorenko, T. (2012). *Evaluating the Validity and Reliability of the Diversity Icebreaker Questionnaire.* partial fulfillment of the requirements for MBA degree, Humboldt-Universitat zu Berlin, School of Business and Economics.

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin, & S. Worchel (Eds.), *The social psychology of intergroup relations* (pp. 33-47). Monterey, CA: Brooks/Cole.

Tenopyr, M. L. (1988). Artifactual reliability of forced-choice scales. *Journal of Applied Psychology*(73), 749-751.

Trompenaars, F. (1993). *Riding the waves of culture: Understanding cultural diversity in business.* Avon: The Bath Press.

VanGundy, A. B. (1981). *Techniques of structured problem solving.* New York: Van Nostrand Reinhold.

Vittersø, J. (1999). Subjective well-being versus self-actualization: using the flow-simplex to promote a conceptual clarification of subjective quality of life. In A. Michalos, *Social Indicators and Quality of Life Research Methods: Methodological Developments and Issues.* Dordrecht: Kluwer Academic Publishers.

Wallach, M. A., & Kogan, N. (1965). *Modes of thinking in young children: A study of the creativity-intelligence distinction.* New York: Holt, Reinhart, & Winston.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measure of positive and negative affect: The PANAS scale. *Journal of Personality and Social Psychology, 54*, 1063-1070.

Weiner, I. B., & Craighead, W. E. (Eds.). (2010). *The Corsini Encyclopedia of Psychology* (4th ed., Vol. 2). New Jersey: John Wiley & Sons.

Whorf, B. (1956). *Language, Thought & Reality.* Cambridge, MA: MIT Press.

Wilden, A. (1980). *System and structure: Essays in communication and exchange.* London: Routledge.

Yamagishi, T. (1988). The provision of a sanctioning system in the United States and Japan. *Social Psychology Quarterly, 51*, 256-271.

Yukl, G. (2001). *Leadership in Organizations.* New Jersey: Prentice Hall.

Zhang, L.-F., & Sternberg, R. J. (2006). *The Nature of Intellectual Styles.* New Jersey: Lawrence Erlbaum Publ.