



KATHOLIEKE
UNIVERSITEIT
LEUVEN

DEPARTEMENT TOEGEPASTE ECONOMISCHE WETENSCHAPPEN

RESEARCH REPORT 0257

**DOES ONE NEED TO GO A LONG WAY TO DIG
DEEP? AN EMPIRICAL COMPARISON OF
ONLINE AND TRADITIONAL FOCUS GROUPS**

by

**S. DEWITTE
H. HENDRICKS
M. JANSSEN**

D/2002/2376/57

Does one need to go a long way to dig deep?

An empirical comparison of online and traditional focus groups

Siegfried Dewitte

University of Leuven

Hendrik Hendricks

University of Gent & Censydiam

Madeleine Janssen

Censydiam

Running head: Online focus groups

Author notes

Correspondence concerning the article should be addressed to Siegfried Dewitte, Department of Applied Economics, University of Leuven, Naamsestraat 69, 3000 Leuven (Belgium), e-mail: siegfried.dewitte@econ.kuleuven.ac.be

Acknowledgements

The first author was supported by a grant from the Fund of Scientific Research, Flanders (Belgium). We thank Marnik Dekimpe for his insightful comments on an earlier version of this paper. We thank the students of MSM 2001-2002 for their invaluable help with data collection.

Does one need to go a long way to dig deep?

An empirical comparison of online and traditional focus groups

Abstract

We evaluate the potential of online focus groups to trigger deep level customer information, which is the major aim of focus groups. We do so by comparing its yield to that of its traditional counterpart. The traditional focus group substantially outperforms online focus groups in level of disclosure, in number of words generated, and in number of ideas generated. However, participants do not reveal these differences in their self-reports. Further, in line with the view that disclosure requires gradual warming up, we find increases in disclosure during the interview in traditional focus groups but not in online focus groups. We conclude that in its present form, the online focus group is not particularly suitable to yield deep level customer information. Starting from our finding that the difference in communication speed explains the differences in disclosure, we suggest some methodological improvements to online focus groups that might increase their yield.

keywords: focus groups; disclosure; group dynamics

Introduction

Focus group research is a popular device to collect market information (Krueger 1994). Out of about the €18 billion (€1 ≈ \$1) spent on market research in 2001 worldwide (Esomar, 2002), roughly 15% (€2.5 billion euros) were spent on focus group research (Amarach, 2001). In focus groups, people gather in small groups ($n = 4$ to 12) and discuss a certain topic under the active guidance of a moderator. In comparison with quantitative research (e.g. surveys), focus groups allow deep level motivations and feelings to surface that are normally inhibited by social norms and conventions (Callebaut et al. 1999). Although the cost can be quite high (about €4000 for one session, e.g. Rogers Media, 2002), the type of information collected might be worthwhile. Depth interviews can also bring this type of information to the fore, but focus groups seem to have (at least) one additional advantage: they engage group processes that enhance disclosure (Krueger 1994).

In the last decade, the internet has become a popular way for conducting market research (Gaiser 1997; Mahajan & Venkatesch, 2000; Taylor 2000). Although many advocate the web's potential (Gaiser 1997; McDaniel & Gates 1999; Schneider, Kerwin, Frechtling & Vivari 2002; Walston & Lissitz 2000), internet market research has not been without its problems. Representativeness has been a major concern because internet users tend to have enjoyed a higher level of education than their non-surfing peers. On the other hand, cost reduction has been considered a major advantage. Discussing the pros and cons of internet market research is beyond the scope of this article (for overviews see Chase & Alvarez 2000; Ilieva, Baron, & Healy 2002; McDaniel & Gates 1999; Walton & Lissitz 2000). Rather, we focus on the differences between traditional and online formats of focus groups. We attempt to

answer the question whether online focus groups might replace traditional focus groups, or even outperform them in certain circumstances.

Both in literature and in practice, two major positions can be distinguished. Some believe online discussions reduce social inhibition or self-presentational concerns, and hence facilitate disclosure about sensitive subjects (Chase & Alvarez 2000; Kiesler et al. 1999; van Nuys, 1999; Walston & Lissitz 2000). Therefore, online focus groups might be an excellent tool to reach one of focus groups' major aims: facilitate disclosure of delicate information such as personal beliefs and attitudes. Further, one could derive that people who are naturally less eager to disclose (i.e. shy people) will benefit more from these "private" circumstances than more extrovert people.

In contrast, others consider group processes as a quintessential feature of focus groups, and are convinced that flesh-and-blood presence supports these group processes irreplaceably (Callebaut et al. 1999; Greenbaum, 2001; Krueger 1994). The latter position, in its extreme form, would imply that online focus groups would miss their purpose completely. Further, if group processes enhance disclosure, traditional focus groups should be characterized by an increasing level of disclosure during the session whereas online focus groups should not.

These conflicting stances lead to the following hypotheses:

H_{1a}: Online focus groups lead to more in-depth disclosure than do traditional focus groups and

H_{1b}: This will be more the case for shy people than for extroverted people.
versus

H_{2b}: Traditional focus groups will lead to more in-depth disclosure than do online focus groups and

H_{2b}: This difference will exacerbate during the session.

Note that H_{1a} is in contrast with H_{2a} and H_{2b} . H_{1b} is not.

Previous research

We next review empirical research that (1) is relevant to our research question and (2) that compared both types of focus groups. Sharing the kind of information focus group researchers want to get out has a certain cost for participants: It might be embarrassing, or may be used against the person at a later stage. One consequence is that self-presentational concerns may be quite high in focus groups (Wooten & Reed, 2000). Another consequence might be that the other members might consider actually conveying information as an act of cooperation. In general, people's behavior is heavily influenced by the dictum of reciprocity (Cialdini, 2001; Wedekind & Milinski, 2000). If people share something with you (whether or not you need it), you feel obliged to return something. Combining both insights might help understand the success of focus groups: One disclosure brings along the other. Roberts and Sherratt (1998) showed that the strategy in which a player adds slightly to an opponent's offer leads to high and robust levels of cooperation.

Focus groups might rely on a comparable escalating form of cooperation. If one person begins disclosing about her inner self, the others may be inclined to follow. In this way, the group digs deeper and deeper. The relevant question is whether this snowball effect can also be triggered in online group discussions. There is evidence that anonymity might hinder rather than enhance such processes (which would be in line with H_2). De Cremer, Snyder, & Dewitte (2001) showed that the effect of trust on cooperation is higher when accountability to others is high than when it is low. Applied to focus group research, the relative anonymity of online settings might prevent trust to increase gradually. However, Moon (2000) showed that people are more willing to disclose to a computer when the computer first reveals something

about itself than when it does not. Moon attributed this finding to the basic human tendency to reciprocate discussed above. It is not clear whether interacting with a computer is comparable to interacting with others by means of a computer.

Further, some researchers compared both formats directly. Chase and Alvarez (2000) reported their experience with online and traditional focus groups. These groups were not set up as a controlled experiment, but were part of different research projects. They noticed that online discussions are more volatile and require more interventions on part of the moderator (i.e. she needs to be highly skilled in electronic communication). Therefore, in the present research, we will use the same discussion guides in both formats and train the interviewers for their specific format. Further, Walston and Lissitz (2000) set up an experiment to explore the differences between both formats. They found that face-to-face interviewees were more likely (1) to have felt embarrassed to reveal something about themselves, (2) to have experienced the facilitating group effect, and (3) to have been concerned about what the moderator (but not the other participants) thought about them, although the effects were not large. The first and third finding are in line with H_1 , whereas the second is rather in line with H_2 . A potential problem was that the participants were not randomly assigned to the groups. The sample of online participants might be extremely computer literate, and hence reduce the potential negative features of online focus groups. Potentially more damaging, the groups were not of equal size. In fact, the online focus groups were larger than the traditional focus groups (as they are in practice too, Walston & Lissitz, 2000). Fern (1982) showed that group size matters in group discussions. He showed that the return of larger groups decreases with increasing numbers of participants (in terms of number of ideas per individual). Further, he showed that groups were less productive than individuals. Therefore,

larger groups might damage online focus groups. To reduce this concern, in the present study participants were first recruited and then randomly assigned to one of the formats. Further, the group size was kept constant.

Finally, Schneider et al. (2002) tested several hypotheses on the difference between both formats. They found that online focus groups led to more individual comments but to much fewer words generated (both in total and per minute). Brief expressions of (dis)agreement were also more likely in online focus groups than in traditional focus groups. Further, the authors showed that contributions were more egalitarian in online focus groups (i.e. the variability of words generated was smaller). They attributed this to the lack of status information in online focus groups, which might inhibit people who feel slightly inferior and hence contribute less. There is an alternative explanation: If word generation resembles an exponential distribution, the lower variance in online focus groups might be due to the lower frequency of words generated and not to individual differences. Further, like Walston and Lissitz (2000), Schneider et al. (2002) did not randomly assign participants to conditions, nor did they keep the groups size constant (see above for comments on that). The obvious strength of Schneider et al.'s (2002) study was their use of behavioral measures. Their data showed clearly that online focus groups are severely constrained by word generation speed, a characteristic that did not show up that strongly in previous research (see above). This might have implications for the hypotheses tested in this study. That is, if group processes rely on slowly escalating goodwill that entails disclosure with it (see H₂), then disclosure (number of words uttered and breadth of ideas) should be highly related to depth of disclosure. Specifically, digging deep should imply going a long way. On the other hand, according to the first hypothesis,

online interviewing might obviate this road to deep disclosure because of lower social inhibition. In sum,

H_{1c} : The correlation between word count and depth of disclosure should be smaller for online focus groups than for traditional focus groups.

H_{2c} : The effect of focus group type on depth of disclosure should be mediated by breadth of disclosure. In other words, one needs to go a long way to dig deep.

Note that statistically, H_{1c} does not depend on H_{1a} and H_{1b} , whereas H_{2c} can only be tested (but is not implied) if H_{2a} holds.

Further, we will critically test the hypothesis that online focus groups are more egalitarian (Schneider et al., 2002). Although we expect, in line with previous research, that the variability of the word count measures will be larger in traditional than in online focus groups, we suspect that this is due solely to the underlying exponential process (higher frequencies are accompanied with higher variances). The following hypothesis can be derived:

H_{3a} : Variance of the raw frequencies is higher in traditional than in online focus groups, but this difference disappears for log-transformed frequencies.

H_{3b} : (dependent on H_{2a}) The difference between both formats relies on non-overlapping series rather than on a minority of extremely high scores in the traditional focus groups.

A final (explorative) hypothesis deals with the probing of interviewers. We will examine whether the formats differ in terms of probing rates, and whether probing enhances disclosure.

The present study

In the present study, we experimentally manipulated the format of focus group. Eight unisex focus groups of four participants each were run, four online and four in a

traditional format. The participants were first recruited and then randomly assigned to one of the groups. They did not know each other and the groups were of equal size. The same discussion guide was used in both formats. The moderators were students enrolled in an advanced marketing program. Each moderator conducted two groups. The most important measures were behavioral and self-reported disclosure. In addition, we measured the number of interventions of the moderators.

Method

Participants

During their lectures, we invited students of several study orientations to participate in a group discussion. To reduce background differences, we selected students that had at least two years of experience as a university student. They received €10 in return for their participation (€1 ≈ \$1). The participants were randomly assigned to the online or traditional focus group condition. They were assigned to unisex groups of five people with the restriction that all participants had to be complete strangers to each other. They were then invited by e-mail. Although the groups consisted of only 4 people, we invited 5 people to reduce variability in group size due to absence. The fifth person participated in another study for the same fee. In total, 31 people participated. In one (traditional male) group, only 3 people showed up. Sixteen were women and 15 men. The age ranged between 20 and 25 years ($M = 22.2$). Note that all but one participant had chat experience.

Discussion guide and interviewer selection

As in practice (Krueger 1994), we first constructed a discussion guide that was used invariably in the 8 focus groups. It was constructed by a group of 5 peer students (including the moderators). It started with a couple of warming up questions. There

were 9 real questions. The sensitivity of the questions increased from 1 to 3. From the third question on, the level of sensitivity was of a comparable level throughout the interview (The discussion guide can be obtained from the first author) The general topic was physical and public appearance, topics that were considered to be sufficiently sensitive to allow social inhibition effects to come into play and to be relevant to the majority of people.

Further, an experienced interviewer trained the interviewers. They conducted two exercise sessions with the same discussion format in advance. We preferred to work with inexperienced interviewers because skilled online interviewers are not easily found. Therefore, using experienced interviewers might confound focus group type with interviewer experience and damage internal validity (i.e. differences between focus group types might be ambiguous).

Procedure

Traditional focus groups

The students were led to a small room where there were five chairs and a table. There was a table with refreshments, coffee, and cookies. The windows were made intransparent to avoid distraction from outside. On the other side of the room, there was a digital camera filming the participants. One person handled the camera and left after a couple of minutes. There was also a tape-recorder that the interviewer started herself.

The interviewer kindly welcomed them and invited them to take a seat. The interviewer introduced herself, sketched the nature of focus groups and invited them to take a drink whenever they wanted to. She also announced that she would tape-record and film what they were going to say. She assured the participants that they

would only use the tapes to type out the transcripts and that their names would be removed. Then she started the warming up questions of the interview.

Online focus groups

The students were led to a pc-room of appr. 30 pcs by an independent guide. In the room, the interviewer was sitting in the back of the room and some students they did not know ($n= 4-5$) were filling the room to avoid that the participants could spot their discussion partners. Several seats separated them. They logged in a controlled chat room. The interviewer introduced herself online and started the interview. The guide stayed in the room to help in case of technical problems.

Software

The software (Blackboard ®) allowed participants to type simultaneously. The typed phrase was published on the common panel only after the enter button was pushed. This allows for simultaneous threads to evolve and might compensate for the fact that people type slower than they talk (Schneider et al. 2002). Before each line of input, the name of the contributor was mentioned.

Final Questionnaire

After the last interview question, the participants were asked to complete a short questionnaire. We measured their self-declared extroversion (one item), the number of ideas the group generated, the level of personal relevance of their answers, their honesty, and their level of comfort with the interviewer, the setting, English, typing, and the topic. Then participants were thanked and paid. The whole experiment lasted about 1.5 hour (with 1 hour and 10 minutes effective for the interview). One week later, they received an e-mail that explained the purpose and results of the study.

Results

The interviews of the four traditional focus groups were transcribed verbatim. We first report preliminary analyses and then test the hypotheses introduced above.

Sample checks

We conducted ANOVAs with self-declared extroversion and level of comfort (4 items, Cronbach's $\alpha = 0.73$) as dependent variables, and focus group type and gender as independent variables and session as a disturbing variable nested in focus group type. Neither level of felt comfort (online: $M = 4.06$, trad.: $M = 3.90$) nor self-declared extroversion (online: $M = 3.25$, trad.: $M = 3.13$), differed as a function of focus group types, all $F_{s(1,4)} < 0.65$. There were differences between sessions for level of comfort ($F(4,23) = 4.10$, $p < .02$) but these differences were not related to focus group type. In sum, the samples were comparable on the relevant dimensions.

Measures of depth and breadth

First, we counted the words in the scripts for every person and every question. Because high means were related to high variances, we log transformed the counts to stabilize variance. Further, questions differed with respect to the length of responses they evoked. Therefore, the log-transformed scores were z-standardized per question.

Second, all scripts were read three times by students comparable to those of the sample: once by a person who read them all (the anchor), once by the interviewer herself, and once by another interviewer. The three coders scored the responses to the individual answers person by person concerning breadth (number of ideas conveyed) and depth (level of disclosure) separately. To safeguard comparability, we compared every interviewer with the anchor and adjusted accordingly. Further, we removed codings if their correlation with the two other codings was too low.

Factor analyses on the 9 questions revealed one large factor explaining about half of the variance, and one smaller one with eigenvalues slightly above 1 (for word count, breadth, and depth). Because all questions loaded on one factor, we considered only the general factor for each measure of disclosure in further analyses (Cronbach's alphas for word count: 0.87, for breadth, 0.65, for depth: 0.86).

So, in total, we have five measures of disclosure: word count, coded breadth, coded depth (the latter two by combining codes of 2-3 converging coders), self-reported breadth, and self-declared depth (the latter two on a five-point scale with high numbers reflecting high disclosure). Table 1 shows the relations between these variables (above diagonal). Strikingly, it shows that self-report measures are not related to the objectively observed variables, whereas the relations among the observation measures were very high.

Table 1. Bivariate correlations (n = 31) among the measures of disclosure

		1	2	3	4	5
Depth	Coded	1	14	86	81	07
	Self-reported	2		14	15	24
Breadth	Word Count	3	<i>.62 / .64</i>		79	20
	Coded	4	<i>.85 / .88</i>	<i>.83 / .76</i>		04
	Self-reported	5				

^a Correlations in bold are significant at the .0001 level, those in italics at the .01 level.

^b Above the diagonal are the correlations in the whole sample. Below the median are the correlations within (respectively) the online (n = 16) and traditional focus groups (n = 15).

Overall disclosure as a function of focus group type (H_{1a} and H_{2a})*Observed variables*

For word count, depth, and breadth, a separate repeated measures ANOVA was conducted with the level of disclosure for the nine questions as repeated measures, focus group type and gender as independent variables, and session nested within focus group type (Note that possible interviewer effects are controlled for because they are subsumed under session).

Table 2. Word count, observed breadth of disclosure, and observed depth of disclosure as a function of Focus group type, session, and gender.

FGT	Gender	Ses ¹	n	Word count			Breadth		Depth	
				Raw Mean	Trans. Mean	Sd	Mean	Sd	Mean	Sd
Trad.	Women	1 (m)	4	1424	0.53	0.4	3.78	0.7	4.80	0.8
		8 (v)	4	1899	0.77	0.2	4.42	0.2	5.26	0.2
	Men	4 (v)	3	2147	0.73	0.3	4.31	0.4	4.69	0.1
		5 (m)	4	1586	0.47	0.4	3.84	0.6	4.21	0.7
Overall				1738	0.62	0.3	4.07	0.5	4.74	0.6
Online	Women	2 (t)	4	476	-0.77	0.3	3.65	0.4	3.61	0.6
		3 (a)	4	418	-0.80	0.1	3.19	0.1	3.08	0.2
	Men	6 (t)	4	627	-0.41	0.4	3.57	0.4	2.93	0.3
		7 (a)	4	687	-0.36	0.6	3.70	1.0	3.60	0.9
Overall				552	-0.58	0.4	3.53	0.5	3.30	0.6

¹ note. Ses. = Session. Within parentheses is the interviewer. Note that the conclusions do not change when interviewer is included in the analyses instead of session and gender.

Table 2 shows the means (across questions) as a function of Focus group type, gender, and session. For the analyses of word count, we used the standardized log transformed values in all analyses. In addition, Table 2 also shows the raw word count measures to give an impression of the size of the effect.

Table 3 shows the summary statistics of the ANOVAs (between-subject analyses) for these variables. We found strong main effects (controlled for session differences) of focus group type for word count and depth of disclosure and a marginal effect for breadth. Participants in the traditional focus groups talked more than three times as much as wrote the participants in the online focus groups in about the same time. The scripts were also judged to contain more ideas, although this difference was smaller. Further, depth of ideas conveyed differed greatly.

Table 3. The ANOVA results for the between subjects factors (focus group types, gender, and session).

	DoF*	Number of words F	Breadth of ideas F	Depth of ideas F	Depth of ideas/self-report F
Focus group type	1,4	191.89 (p<.0001)	5.83 (p=.07)	27.45 (p<.01)	0.02 ns
Gender	1,23	1.65 ns	0.25 ns	2.56 (p=.12)	0.02 ns
Focus group type * gender	1,23	2.75 (p=.11)	0.39 ns	1.45 ns	11.11 (p<.002)
Session (nested in gender*focus group)	4,23	0.42 ns	1.39 ns	1.80 (p=.16)	0.64 ns

* DoF = Degrees of Freedom.

To summarize, we found that traditional focus groups led to more disclosure than did online focus groups, which supports H₂ and is at odds with H₁.

Self-report variables

Two similar ANOVAs were conducted with focus group type and gender as independent variables, and session nested within focus group type: One with self-reported breadth of ideas and one with self-reported depth of disclosure as dependent variables. The measure reflecting number of ideas (breadth – self-report) showed no effects (all $F_s < 1.06$). Depth of disclosure (self-report) was not affected by focus group type in itself, but by a significant interaction between gender and focus group type (see Table 3, last column). Men reported having disclosed more deeply online ($M = 4.13$) than in traditional focus groups ($M = 3.50$), whereas women reported the opposite (online: $M = 3.57$, traditional: $M = 4.13$). In fact, the objective data show that all participants disclosed more in traditional than in online focus groups.

Further, for number of words, depth, and breadth, the correlations with extroversion were lower in the online focus group (respective $r_s = -0.34$ ($p > .20$), $-.06$ and -0.24 , ns.) than in the traditional focus groups (respective $r_s = 0.14$, 0.10 , and 0.10 , ns.), which was predicted by hypothesis H_{1c} . However, the expected interaction between extroversion and focus group type was far from significant (all $F_s(1,23) < 1.28$)

Dynamics of disclosure: Does disclosure enhance more during the traditional focus groups than during online focus groups? (H_{2b})

To test the hypothesis, we looked into the evolution of disclosure. To reduce the influence of individual questions, we split the nine questions in three groups of three: the first, the second, and the last three. If there is a differential effect, then the gap between both focus group types should widen with time past. The repeated measures ANOVAs we conducted above were redone with three repeated measures (the three parts). To explore dynamics, we focus on the interaction between position (1, 2, or 3) in the series and focus group type. Indeed, for depth ($F(2,22) = 4.48$, $p < .02$) and

breadth of ideas ($F(2,22) = 7.80, p < .003$), the interactions were significant in the expected direction (see Figure 1 for the trend plots). For number of words, the interaction was similar but only marginally significant: $F(2,22) = 3.28, p < .06$ (for online focus groups: 182, 195, 175, respectively, for traditional focus groups: 578, 438, 722, respectively).

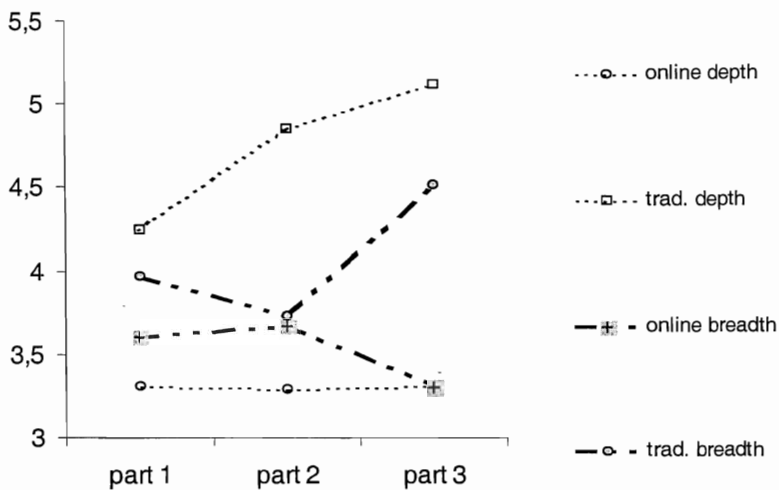


Figure 1. The evolution of disclosure (breadth and depth) during the interview as a function of focus group type.

In sum, the analysis supports H_{2b} . In contrast with online focus groups, traditional focus groups can be characterized by an increase in disclosure in the course of the session.

Does one need to go a long way to dig deep (H_{1c} & H_{2d})?

Table 1 already showed that breadth and depth are related very strongly in this study. We further looked at the correlation between the three measures of disclosure within both conditions. H_{1c} predicts that this relation should be lower in online focus

groups because online communication may obviate the necessity to gradually pave the way towards self-disclosure because of its impersonal character. Table 1 (below the diagonal) showed that there is no difference in strength of correlations between the measures of disclosure in both conditions. In sum, H_{1c} is not supported.

To test H_{2d} , we used a mediation analysis. If the quantity is a necessary condition to yield quality or depth of disclosure, then the effect for focus group type on depth of disclosure should be mediated by word count. Mediation requires (Baron & Kenny, 1986) a main effect of focus group type on both word count and depth of disclosure, and a relation between word count and depth of disclosure. Both condition have been met (see Tables 2&1). The final requirement is that the effect of focus group on depth of disclosure should disappear when word count is included as a covariate. Table 5 shows the F-value for the effect of focus group (column 1) and of word count (column 2) on depth of disclosure, and the effect of focus group when word count is included as a covariate (column 3). We show this for the three parts of the interview. The effect of focus group type on depth of disclosure disappears when word count is included in the analysis. The data support H_{2d} .

Table 5. F-values illustrating the mediating role of word count in the relation between focus group type and depth of disclosure.

	From FG type to depth (F-value)*	From word count to depth (F-value)	From FG type to depth via word count (F-value)
Part 1	28.52	42.57	0.60**
Part 2	33.58	11.92	3.31*
Part 3	71.69	14.68	0.61**

All $F_{s(1,27)}$ are significant at the .003 level or better except for * : $p = .07$ and **, $p > .40$.

Are online focus groups more egalitarian?

We tested the hypothesis that online focus groups lead to more egalitarian contributions (Schneider et al. 2002) by comparing the variances. Although the variance of the raw word counts was larger in the traditional ($S^2 = 284,067.1$) than in the online focus group condition ($S^2=60,905.63$, $F(15,16) = 4.664$, $p < .01$), replicating Schneider et al.'s finding, this difference disappeared after the log transformation. The variance of number of words for online focus group participants ($S^2 = 0.18$) was even larger than that in the traditional focus groups ($S^2 = 0.10$, but not significantly so, $F(16,15) = 1.8$, $p > .05$). This supports H_{3a} and suggests that the difference in variances is an artifact due to differences in frequencies.

To test H_{3b} , we looked at the extreme scores in both groups. The standardized (log transformed) scores of the lowest four scorers (on word count) in the traditional focus groups are 0.03, 0.10, 0.13, and 0.50. Compare these to the highest four scorers in the online focus groups: 0.49, 0.00, -0.16, and -0.25. This shows that the most disclosing online participant outperformed only three of the most reticent participants in the traditional focus groups on words generated. This shows that the higher average in traditional focus groups is not due to a few dominant contributors, but to the vast majority of the contributors.

The role of probing.

Further, we looked at the number of probes the interviewers gave. For all scripts, the number of probes was counted per question and per participant. When a question was addressed to all of the participants (or whether it was not clear to whom it was addressed), it was counted for all participants. We conducted a hierarchical ANOVA, with number of probes as dependent variables (averaged over questions) and focus group type and gender as independent variables, and interviewer nested in focus group

type. The analysis revealed significant differences for interviewer: $F(2,25) = 19.9, p < .0001$. One interviewer (in the online condition) probed her interviewees about 84 times on average, whereas the three others ranged between 48 and 57. This difference also explains that the main effect for focus group (online: $M = 69.4$ vs. traditional: $M = 51.9$) was not significant: $F(1,2) = 1.05$. The most intriguing difference, however, was that for gender: $F(1,25) = 53.3, p < .0001$. Men triggered more probes ($M=74.1$) than did women ($M=48.6$). This finding is in line with the experience all of the interviewers expressed that the male focus groups were harder to conduct. This might have to do with the topic. Men might be less used to talk about physical appearance.

Interestingly, correlation analyses showed no significant relations between number of probes and level of disclosure (five measures). Only breadth of ideas was related slightly to number of probes: $r = -0.35, p = .05$. However, this is probably an artifact because one interviewer in the online session (with lower levels of breadth) probed more than the others. Indeed, the relations were close to zero in both conditions separately (online: $r = 0.06$ and trad.: $r = -0.19$).

In order to check whether number of probes could explain some of the results reported above (especially the null effect of gender), we repeated all ANOVAs reported above with number of probes as a covariate. First, number of probes had no effect on any measure of disclosure (all $F(1,22) < 1.24, ns$). Further, the other effects, including the null effect of gender, remained unaffected by the inclusion of this covariate. This suggests that although men were probed more often, this did not hide a gender difference in disclosure. Further, it adds to our previous findings that the effect of focus group type was not driven by interviewer differences.

Discussion

The data do not leave much doubt that the results reported in this study support the position expressed by H₂ to a larger extent than the position expressed by H₁. First, traditional focus groups do not only lead to many more words generated than do online focus groups, they also lead to higher levels of breadth and depth of disclosure. Further, the gap between traditional and online focus groups deepens as the session proceeds, which is in line with the group dynamical view expressed in hypothesis 2. In addition, breadth and depth are closely related, and the strength of this relation is not moderated by focus group format. Related to this, number of words generated mediated the effect of focus group type on depth of disclosure. Finally, shy people do not benefit substantially more from online environments than do extroverted people, which was predicted by H₁.

Our results suggest that online focus groups face severe problems in bringing deep information to the surface. They lead to substantially lower levels of deep disclosure. In the present sample, this huge difference cannot be overcome by the lower social inhibition in online settings (Wooten & Reed, 2002). Rather, we present evidence that the social context in traditional focus groups works as assumed: disclosure increases as (discussing) time passes, which is in line with the general idea that disclosure (as a form of (informational) cooperation among individuals) requires warming up (Moon 2000; Roberts & Sherratt; 1998). Apparently, one needs to go a long way to dig deep. In the remainder, we first attempt to interpret our findings in terms of underlying processes, then address some interesting details of our results, discuss some possible limitations to our methodology, and then suggest some possible enhancements to online focus groups methodology.

Interpretation of the findings

Although there are slight session and interview differences (see Table 2), we observe that the lion's share (68%) of the variance in disclosure can be explained by focus group type. Live contact, being the most obvious difference between both formats, possibly leads to mutual trust that gradually increases and leads to deeper self-disclosure. We find clear evidence for this process, and show that increasing depth is mediated by number of words generated. If this interpretation holds, then online focus groups are doomed to remain fairly limited in their use, unless virtual focus groups become widely available (Chase & Alvarez 2000).

However, there might be another but related explanation that fits as nicely with our data and is inspired by the huge difference in number of words generated. In this view, the major driver of the differences is a purely technical difference between both formats that interacts with a human discourse characteristic. Specifically, people talk faster than they can type. If we further assume that human conversation speed is fine-tuned on talking speed and not on typing speed, online focus groups might suffer from understimulation (which might be further exacerbated by the lack of other stimuli such as non-verbal and paravocal cues). This difference has not gone unnoticed in literature (Schneider et al., 2002; White, 2000) but has not been considered as a major problem hitherto. However, in our opinion, this difference is not innocent and might (at least partially) drive the huge differences between both formats. First, number of words generated mediates the effect of focus group type on depth of disclosure. Further, the paucity of the information generated in the early phases of the online focus groups might prohibit the escalation of information exchange that is probably crucial to dig under the surface. Online participants' contributions remain fairly

constant, whereas traditional focus group participants become more and more involved (see Figure 1).

Post-interview discussions with the interviewers and inspection of the scripts confirm that human conversation characteristics might play a role. Interviewers in the online groups found it frustrating to keep track of four dialogues and were concerned that participants would lose focus. The online scripts also look rather scattered (Excerpts can be requested from the first author). Some questions are followed up only after some time, which might frustrate participants or at least cool their enthusiasm to disclose. If one considers the time between contributions (which is much larger due to the slow pace), this might be a real concern. So, online interviewers face a dilemma: elaborate on one interviewee, or keep track of all contributors. Because of the time limitations, they attempted to involve all participants, and in this process, they perhaps traded off depth for equal participation. Notice that this alternative explanation is not at odds with the one expressed in hypothesis 2. It merely adds one typical understimulation problem of online focus groups (speed of typing) to the list of other understimulation problems (lack of non-verbal information, approving nods of the interviewer, attention from the others, etc.). Future research should determine which of these is/are the real culprit(s) that is (are) responsible for the gap between both formats. For instance, if seeing and hearing each other is really required, then (expensive) virtual focus groups might help out (Chase & Alvarez, 2000). If being in each other's presence is required, then even virtual focus groups will not be able to do the job. However, if conversation speed is the major problem, then other changes may do the job in a cheaper way (see further - recommendations). (At this point, it might be useful to note that the use of videoconferencing technology in focus group research pertains to the contact between

the session and the client, and not (to our knowledge) to the focus group participants (McDaniel & Gates, 2001).

Self-report measures of disclosure and interview probing

Had we only collected self-report data, our conclusions would have been drastically different (compare ours with Walston & Lissitz', 2000). The present research dramatically points to the problems of self-report measures in cases where consumers do not have clear anchor points. Interestingly, we found that men thought they were more expressive online whereas women did so in traditional focus groups. In reality, they were all more expressive in the traditional groups. The only thing that is somewhat in congruence with the self-declarations is the fact that men typed slightly more than women (but the interaction is not significant, see Tables 3 and 4). Therefore, we suspect that the self-report has more to do with stereotypes ('men are more computer literate' vs. 'women are more talkative') than with real differences. This finding may also shed another light on some positive findings that rely on self-report measures (e.g. Chase & Alvarez 2000; Tse 1999; Walston & Lissitz 2000). We strongly suggest that future research includes both subjective and objective measures.

Another finding worth mentioning is the fact that probing did not make a real difference. Although interviewers expressed that men were more reluctant to disclose and that they probed them more, our analyses suggested that the probes did not obscure a gender difference in disclosure. Perhaps, men are less used to talk about physical appearance. Therefore, they may need more time to get their ideas sorted out, which may trigger more probes from the interviewers. In the end, however, our analysis suggests that probes had no effect on disclosure.

Limitations of the present study and directions for future research

We briefly discuss possible limitations. First, as we used a one-item measure of extroversion, our failure to find an interaction between extroversion and focus group type should be approached with caution. Second, we acknowledge that our sample's homogeneity (in terms of age computer literacy) prevents us from drawing strong conclusions to the general population. However, we fear that the difference we found among computer literates will only exacerbate in the general population. Third, the discussions were not held in the participants' mother tongue. Before they agreed, they knew that the discussion would be in English, but this factor might still be a confounding factor as writing and talking might be differentially affected by imperfect language proficiency. However, as these students are used to written English (papers, internet) perhaps more than to spoken English, we are confident that proficiency cannot explain away the large gap we found.

Further, the interview duration was rather short, which might prevent disclosure to fully blossom (Callebaut et al. 1999). The fact that we did find this divergence in the short time span makes the findings even more impressive. The short time span has another aspect to it, however. For the sake of obtaining a clean design, we kept the duration constant for both formats. Online focus groups might benefit from dropping the time constraints. The low communication speed in online focus groups might be compensated for by longer interview time (although attention focusing might suffer from that adaptation, see above).

Fifth, the interviews were done by unskilled interviewers. We did so to safeguard internal validity (see above) but we acknowledge that this might damage the generalizability of our results. Interestingly, interview techniques used in online focus groups might be inspired too much by those used in traditional focus groups (not just

in our experiment but in general). Therefore, our data might also point to a need for interview techniques that fit better with online conversation.

Further, the findings probably rely on the (sensitive nature of the) topic. According to our theory, the difference between focus group types should reduce for less sensitive topics. On the other hand, focus groups have been designed for triggering self-disclosure and are better replaced by (online) surveys if one is after neutral information.

Finally, although our self-report measures do not show a difference in comfort with the environment, the online environment might have been less comfortable than the traditional environment. Future research might benefit from allowing people to participate at home while keeping the randomization procedure intact.

So, in general, we remain confident in our results, and think the next steps might be a replication of the findings with the online participants at home, allowing them more time to type their opinions, and tailor interview trainings towards the online format.

Recommendations - Market research implications

We will not reiterate recommended measures to improve online focus groups given by others (see Chase & Alvarez 2000; Curasi 2001; Gaiser 1997; Schneider et al. 2002; Walston & Lissitz 2000). Rather, in our recommendations, we build on our main finding that the number of words differs a lot between both formats and on our interpretation that this may prevent fluent information exchange in online focus groups as currently conceived of. Our recommendations assume that the researcher wants to stimulate self-disclosure. It does not really pertain to the generation of a large quantity of ideas. Further, it ignores the possibility that physical co-presence or face-to-face interaction is required for disclosure to occur.

First, we recommend that the common practice to let more people participate in online focus groups than in traditional focus groups (see Walston & Lissits, 2000, who had one online group of 18 people!) be reversed. Because of the slower generation speed, involvement might be higher when fewer people participate, because this increases the conversation time per participant. An additional advantage is that it will become easier for the interviewer to keep track of the discussion. Further and related to the first recommendation, we suggest that interviewers should allot more answering time to the participants. At this point, it is not sure whether this would lead to higher levels of disclosure. Future research should sort that out.

Third, some simple software adaptations may be in order. Software adaptations may let people see each other typing in real time character by character. This would increase involvement and avoid dead moments. People may guess what the other is typing rather than sit and wait for the line to pop up. The disadvantage may be that overall speed further decreases because simultaneous typing will be less frequent. Showing different windows per participant could circumvent this drawback. The advantage is obviously the increased similarity with speech.

Further, any measure that enhances communication speed might be desirable. For instance, people might use a set of abbreviations for the most common words. The interviewer could offer this lexicon before the interview starts. The question is how far one could go in this: Too many abbreviations might as well ruin the conversation and some people might be better in abbreviating than others. Related to this, the input boxes might be reduced so people have to enter their contribution more often. This would decrease waiting time.

On the part of the interviewer, and in line with Walston and Lissitz' (2000) recommendation that the interviewer uses more than one computer, the interviewer

may benefit from some technical gadgets, for instance different colors for the different participants, word recognition software to increase typing speed, one window for each participant, or warnings when one participant has not been addressed for a certain time.

Finally, as mentioned in the future research directions section, social psychological factors might lower the threshold of intragroup cooperation to take off. For instance, stressing similarity of the participants, increasing group identity, or staging disclosing peer examples (as far as ethical code allows) may all decrease the critical threshold of information exchange and hence enhance cooperation among group members.

Research is needed to evaluate the strength of these measures.

Conclusions

The major contribution of this study is that it showed that online focus groups lead to much slower word generation than do traditional focus groups. This is not interesting in itself. However, it apparently leads to less self-disclosure, not only because there is less time to disclose, but also because the contributions remain below a critical threshold which is required to allow group dynamics to enhance disclosure. Our recommendations to improve online focus groups relied on this generation speed characteristic. Indeed, it appears that one needs to go a long way to dig deep.

References

- Amarach (2001). Tomorrow's news. http://www.amarach.com/news/issue_8.html
- Baron, R.M., & D.A. Kenny (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, 51, 1173-1182.
- Callebaut, Jan, Madeleine Janssens, Diane Op de Beeck, Dirk Lorré, & Hendrik Hendricks (1999). *Motivational marketing research revisited*. Leuven: Garant.
- Chase, L, & J. Alvarez (2000). Internet research: The role of the focus group. *Library and Information Science Research* 22, 357-369.
- Cowley J.C.P. (2000). Strategic qualitative focus group research – define and articulate our skills or we will be replaced by others. *International Journal of Market Research* 42, 17-38.
- Curasi C.F. (2001). A critical exploration of face-to-face interviewing vs. computer-mediated interviewing. *International Journal of Market Research* 43, 361-375.
- Esomar (2002). ESOMAR Annual Study of the market research industry 2001. <http://www.esomar.nl/press/MRIndustry2001.htm>
- Fern, Edward F. (1982). The use of focus groups for idea generation: The effects of group size, acquaintanceship, and moderator on response quantity and quality. *Journal of Marketing Research* 19, 1-13.
- Gaiser, Ted J. (1997). Conducting on-line focus groups. *Social Science Computer Review* 15, 135-144.
- Greenbaum, T (2001). Online focus groups are no substitute for the real thing. *Quirk's Marketing Research Overview*, June 2001.
- Ilieva, J., Baron, S., Healey, N.M. (2002). Online surveys in marketing research: pros and cons. *International Journal in Market Research*, 44, 361-377.

- Krueger Richard A. (1994). *Focus groups. A practical guide for applied research (second edition)*. Thousand Oaks: Sage.
- Mahajan V, & Venkatesh R (2000). Marketing modeling for e-business. *International Journal of research in marketing*, 17, 215-225.
- Richman W.L., Kiesler S., Weisband S., & Drasgow F. (1999). A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. *Journal of Applied Psychology* 84, 754-775.
- Rogers Media (2002). Focusing on focus groups. *Marketing Magazine*, 109, 32.
- Schneider, S.J., Kerwin J., Frechtling J., Vivari B.A. (2002). Characteristics of the discussion in online and face-to-face focus groups. *Social Science Computer Review* 20, 31-42.
- Taylor, H. (2000). Does internet research work? *International Journal of Market Research* 42, 51-63.
- Tse, A.C.B. (1999). Conducting electronic focus group discussions among Chinese respondents. *Journal of the Market Research Society* 41, 407-415.
- Van Nuys (1999). Online focus groups: Market research in web time. *San Jose Business Journal*, November 1999.
- Walston, J.T. & Lissitz R.W. (2000). Computer-mediated focus groups. *Evaluation Review* 24, 457-483.
- Wooten D.B. & Reed A. III (2000). A conceptual overview of the self-presentational concerns and response tendencies of focus group participants. *Journal of Consumer Psychology* 9, 141-153.