

Feedback on 'Coalition for Better Ads' methodology

January 2019

We, entrepreneurs affiliated within IAB Poland – in particular members of AudioVideo Working Group, Adblock Working Group and Research Working Group – express our concern and opposition towards the latest planned 'Coalition for Better Ads' activities.

We would like to stress that the overall goals of the Coalition for Better Ads and the idea behind creating the Better Ads Standards are non-disputable and we strongly support it. We believe that setting good ad standards, backed up by professional market studies, and reflecting the users' expectations, can be the most successful way of addressing the trend of blocking on-line ads. However, we also believe that in order to have a real market impact, the market studies must be conducted in a professional manner and reflect the local market specifics. Thus, after analyzing the documents submitted by CBA regarding the latest stage of the study on ad experience, we state that the project does not meet requirements of reliability and objectivity. Study design is not robust, questions are biased and tested stimuli does not reflect real conditions.

The above mentioned issues raise our great doubts and in our opinion further development and promotion of the study may lead to very adverse effects for the market, including the limitation of its evolution and selective identification of entities operating on it. At the same time, we do not see that priorities of the initiative – the need to protect users and the increasing threat associated with the usage of ad blockers – are appropriately addressed.

In addition, we would like to point out that in the current state of technological development, the majority of market activities focus on optimizing and improving targeting methods that limit the intrusiveness of advertisements by precisely matching them to user's needs and interests. As part of the work of the IAB Poland Working Groups, we conduct a number of initiatives to standardize the market, aimed at improving UX. The quality and convenience of user contact are our priority.

We would like to invite companies united within the IAB to analyze the research methodology and to formulate a common position on this issue. That is why we call for project to be suspended until the appropriate research methodology is established and further actions related to it.

In the further part of the document there are detailed analyzes of methodology of the experimental CBA study, which we share in the interest of the substantive value of research and the shape of the digital communication market. The document includes all doubts, starting from those reported to CBA in 2017 – when the methodology became available to the public – ending with the current ones regarding video online.

[Table of contents](#)

1. Research model tests fake ad experiences.....	3
Research perspective	3
Perspective on the CBA study	4
2. Questionnaire is biased.....	5
Research Perspective	5
Perspective on the CBA study	6
3. Choice of stimuli affects results.....	6
Research Perspective	6
Perspective on the CBA study	8
4. Respondents are not typical web users	8
Research Perspective	8
Perspective on the CBA study	9
5. Summary.....	9
References	10

1. Research model tests fake ad experiences

Research perspective

Online ad experience, in its current state, is based on advanced serving models. Richness of existing solutions lets advertisers and publishers use such tools as content marketing, native advertising, targeting, re-marketing, real-time-marketing, programmatic and many others. Secondly, particular ad formats perform differently depending on type of webpage content and publishers use various models to serve particular formats. Thirdly, quality of contact with an ad very often depends on specific relation between brand (or brand category) and customer. In many cases advertised brands are well known. As a result, online ad experience is a complex phenomenon – resulting from cognitive, attitudinal and behavioral factors prior to ad exposure. As a result, there are huge demands when constructing research models. In practice, number of factors need to be carefully defined and included to ensure that accurate and valid results are produced.

An example of how one of the above-mentioned factors can affect ad experience can be results of IAB Poland/MEC WebTrack study¹. It proves that level of annoyance towards ads depends on relation to advertised brand and even the category itself. Brand users and users of category significantly less frequently perceive exposed ads as annoying in comparison to non-users. As the above mentioned study shows, basing on typical real campaigns, 75% of web users exposed to ads are category users. It lets conclude that testing fake brands can significantly increase level of annoyance toward ads.

ANNOYANCE LEVELS DEPENDING ON RELATION TO BRAND/CATEGORY

To what extent the following statements suit the ad? – It is annoying

	Relation to advertised brand/category		
	Brand users	Category users (brand non-users)	Non-users
Definitely does not suit	22%+	20%	20%
Rather does not suit	24%+	24%+	21%-
Neither suits nor does not suit	34%	35%	35%
Rather suits	12%	13%	12%
Definitely suits	8%-	8%-	12%+
Top 2 boxes:	45%+	43%	41%-
Bottom 2 boxes:	21%-	21%	24%+
Mean:	2,63	2,67	2,74
Sample size:	6,695	16,372	5,584

+/- Significant difference on 95% level, tested vs total.

Source: IAB Poland/MEC WebTrack, 2014, 20 various campaigns, 158 ads.

Another example of how independent variables can impact perception of an ad comes with the comparison of placements. As WebTrack study shows there are strong differences in perception of annoyance towards the same ads but depending on type of service where ad experience was observed. Tested campaigns were run on various types of websites: information, gossip, entertainment, social media etc. In both analyzed cases the same ad was displayed to web users in the natural environment

¹ See an example report from the first wave (2012) of the study:
<https://www.iab.org.pl/wp-content/uploads/2014/01/WebTrack.pdf>

and during their typical online sessions. Both cases prove that testing ads depends on the context and conducting research in the artificial environment may strongly skew the results on ad experience.

ANNOYANCE LEVELS DEPENDING ON PLACEMENT

To what extent the following statements suit the ad? – It is annoying

	Double Billboard (750x200)			Triple Billboard (750x300)		
	Gossip Portal	News Portal	Social Media (SNS)	Gossip Portal	News Portal A	News Portal B
Definitely does not suit	14%-	20%+	16%	17%	20%	27%+
Rather does not suit	22%	23%	27%	24%	29%	28%
Neither suits nor does not suit	41%	39%	41%	42%+	31%	29%
Rather suits	14%	12%	11%	11%	13%	10%
Definitely suits	9%+	6%	5%	6%	6%	6%
Top 2 boxes	37%-	43%	42%	42%-	50%	55%+
Bottom 2 boxes	23%+	18%	16%-	17%	19%	16%
Mean:	2,81	2,62	2,63	2,64	2,54	2,39
Sample size:	856	789	313	366	333	197

+/- Significant difference on 95% level, tested vs total.

Source: IAB Poland/MEC WebTrack, 2014, 2 various campaigns, 2xBB and 3xBB.

Other IAB Poland studies (e.g. TV+WWW=Better Together²) also confirm different attitudes towards advertising depending on the type of service and at the same time other user expectations. The video ad experience on VoD websites is different, another on information portals, and yet another on entertainment websites.

Perspective on the CBA study

Unfortunately, CBA experimental study is carried out in artificial environment (it is based on fake webpages³, fake ads, fake content experience) and tries to describe real-world⁴. As scientific research literature proves⁵, such studies try to ‘generalize results beyond artificial conditions’ while their ‘findings hold true only in the experimental situation – and nowhere else’. One of weak points of the CBA study is lack of external validity (based on real ads, real campaigns, real sites and real UX). It is unknown to what extent the tested conditions depart from everyday reality in key aspects. These key aspects have not even been defined, apart from the statement of replication of the “real world”, while the above analysis based on the IAB Poland/MEC study shows clearly that they cannot be ignored in the development of ad standards. Results of the CBA study do not show how real ad experiences would be shaped in the real media content to which individuals would be exposed in real life. The real context and real User Experience are vital to understand the way members of audience ordinarily experience ads under natural viewing circumstances. CBA methodology isolates too many specific factors which influence outcome.

² 2011, 2014, 2018.

³ Coalition for Better Ads [CBA], A consumer-Centric Methodology: ‘simulated content pages’.

⁴ CBA, op. cit., ‘Coalition for Better Ads uses a survey model that replicates the “real world” experiences of users’.

⁵ B. Gunter, 2000.

2. Questionnaire is biased

Research Perspective

Quality of the research tools translates directly into the quality and credibility of the resulting data. Wording of a question as well as labeling and the order of answers affect the perception of the question by the respondent. It is also extremely important if the scale is balanced (has a mid-point, e.g. 'neither annoying nor non-annoying') or not. On unbalanced rating scales, the respondents are forced to express an opinion because a 'no opinion' option is not provided⁶. In fact, it may force the respondent to give the answer expected by the researcher⁷.

Extremely difficult problems in the case of using unbalanced scales appear when a large part of the respondents may not have an opinion. In the case of testing the level of irritation towards advertising – according to the IAB Poland/MEC WebTrack study – the percentage of such observations reaches 35%. This means that the removal of mid-point of the scale compels more than one-third of the respondents to give a value-based answer. The problem escalates when customers who do not know the brand are taken into account. They significantly more frequently consider ads of unknown brands as annoying. As a consequence, lack of mid-point forces negative answers to a large degree.

ANNOYANCE LEVELS DEPENDING ON BRAND AWARENESS

To what extent the following statements suit the ad? – It is annoying

	Aided Brand Awareness	
	Brand aware customers	Brand unaware customers
Definitely does not suit	19%-	21%+
Rather does not suit	24%+	20%-
Neither suits nor does not suit	37%+	34%-
Rather suits	12%	13%
Definitely suits	8%-	12%+
Bottom 2 boxes	43%	42%
Top 2 boxes	20%-	25%+
Mean:	2,66	2,74
Sample size:	17,871	12,583

+/- Significant difference on 95% level, tested vs total.

Source: IAB Poland/MEC WebTrack, 2014, 20 various campaigns, 158 ads.

The issue of balance also concerns the content of the question itself⁸. In many cases the question has a logical 'middle alternative' or is formulated in a neutral way not to present only the affirmative side of the issue (e.g. 'Do you consider the ad annoying or non-annoying?'). It is important to pay a special attention when using unbalanced scales or unbalanced questions while designing the questionnaire because the results may have a large skew.

⁶ N.K. Malthora, 2006.

⁷ T.N. Judice, S.L. Neuberg, 1998.

⁸ H. Schuman, S. Presser, 1981.

Perspective on the CBA study

In the first part of CBA survey questionnaire there appear balanced scales. Questions about overall consideration online advertising as a good thing⁹ and satisfaction with the overall experience on the web page¹⁰ are unbalanced but scales are balanced and include mid-point. The number of positive and negative answers is equal and a neutral answer option is available. This part of the questionnaire seems to be neutral, the questions are not biased, and the extreme answers are not forced.

The next part of the research tool raises serious doubts, especially two questions having key impact on the results of the study:

How ANNOYING was the ad?

- Not at all annoying
- Slightly annoying [NEGATIVE LABEL]
- Moderately annoying [NEGATIVE LABEL]
- Very annoying [NEGATIVE LABEL]
- Extremely annoying [NEGATIVE LABEL]

To what extent does each of the following words or phrases DESCRIBE THE AD? - Distracting

- Not at all
- A little [NEGATIVE LABEL]
- A moderate amount [NEGATIVE LABEL]
- A lot [NEGATIVE LABEL]
- A great deal [NEGATIVE LABEL]

In each of them both unbalanced question wording and unbalanced scale were used. This method of data collection undermines the credibility of the collected data. Both questions and scales include assumption that ad experience is negative. Even if correlation was observed between particular variables (although its coefficient is unknown), it does not confirm the reliability of the results but only the coherence between questions posed in a very tendentious way. Considering the purpose of the research which is the indication of better ad standards, formulating questions in a biased way leads to self-fulfilling prophecies.

3. Choice of stimuli affects results

Research Perspective

As proven above - the mere fact that ads include known or unknown brands is of enormous importance for the ad experience. However, advertising itself is also a key element of the research composition. The art of creating concepts is a separate branch of marketing, consuming huge budgets, and advertisers devote to their fine-tuning even a few years before the advertisement will see the light of day. The concept communicates specific information about the product and must be closely related to it, which is the subject of separate studies and analyzes¹¹. The following are intrinsic components of the concept: Insight, Benefit and Reason to Believe (RtB). Each of them should be clear, refined and unambiguous. In marketing research, perfect stimuli are not always used, especially at the early stages of the marketer's work. An example may be the so-called story boards, the testing of which aims at the initial assessment of the concept. In such cases, however, research procedures are tailored to the

⁹ 'Overall, I consider online advertising a good thing.' Scale: 'Strongly agree | Agree | Neither agree nor disagree | Disagree | Strongly disagree'.

¹⁰ 'How satisfied were you with the OVERALL EXPERIENCE on the web page?' Scale: 'Very satisfied | Somewhat satisfied | Neither satisfied nor dissatisfied | Somewhat dissatisfied | Very dissatisfied'.

¹¹ E. van Kleef, H.C.M. van Trijp, P. Luning, 2004.

specificity of the measurement – the methodology, the presentation of the materials to the respondents and the analysis. If the three research procedures apply solutions dedicated to the final ads (to which consumers are exposed on a daily basis) the stimuli must bear all the hallmarks of a refined concept.

Example of ads tested in the CBA study in Poland



Source: Consumer Ads Preference Methodology in Poland, December 2017.

Copy quality

The attractiveness of the concept is a key element of ad experience. According to the IAB Poland/MEC WebTrack study, the quality of copies is very strongly correlated with the level of annoyance. Internet users who rate ads better (18% vs. 36% - see: table below), significantly less frequently declare that the ads are annoying for them. Although the quality of the copy is not synonymous with its attractiveness, particular attention should be paid to this aspect in advertising research because the selection of inappropriate stimuli can lead to wrong conclusions.

ANNOYANCE LEVELS DEPENDING ON COPY QUALITY

To what extent the following statements suit the ad? – It is annoying

	Overall evaluation of the ad		
	Low quality ad (bottom 2 boxes)	Mid quality ad (mid-point)	High quality ad (top 2 boxes)
Definitely does not suit	24%+	11%-	28%+
Rather does not suit	19%-	20%-	31%+
Neither suits nor does not suit	21%-	54%+	23%-
Rather suits	17%+	10%-	12%-
Definitely suits	19%+	5%-	6%-
Top 2 boxes:	43%	31%-	59%+
Bottom 2 boxes:	36%+	15%-	18%-
Mean:	2,85	2,78	2,37
Sample size:	8,847	13,495	8,112

+/- Significant difference on 95% level, tested vs total.

Source: IAB Poland/MEC WebTrack, 2014, 20 various campaigns, 20 ads.

Perspective on the CBA study

Copy quality – related to display ads tested previously

Taking into account the requirements of the advertising concept, it should be stated that the advertisements studied so far in the CBA project were of low quality. They had inadequate translation and the three key components of the concept were underdeveloped (Insight, Benefit, RtB). The impact of the quality of these copies on the key results of the study is unknown. Additional research comparing tested ads with natural ones should be conducted. However, it can be assumed that the results towards annoyance levels and ad experience can be strongly skewed. In addition, the fact that fake brands were tested – as it has already been demonstrated above – can have a huge impact on experience. These factors cause that the ranking of formats cannot be considered as credible.

Choice of copies – related to short form video

At this stage quality of tested video ads is unknown. However, the requirements for them as well as the research implications resulting from their selection are the same as in case of the above-mentioned display ads.

Moreover, the selected list of formats is not imprecise and very selective at the same time. It includes a small number of variations of ads. Disproportions are particularly noticeable in the midroll and pod formats. Also, it has not been clarified whether they are ads initiated by the user or autoplay, as well muted or unmuted. In addition, the document contains a proposal to include display formats and other formats based on interaction, what does not seem justified in the context of the study of short video forms. It is also worth noting that the adopted formats are not the most commonly used in Poland, so the application of the results to local realities will be very limited.

4. Respondents are not typical web users

Research Perspective

There are a lot of studies that indicate the strengths and weaknesses of different sampling methods, including panels. Despite many advantages, it should be remembered that panelists are not typical internet users, do not behave and do not answer questions like typical internet users. There are two factors that influence the data gathered via panel. First of all, panel users are the heavy users of the Internet and the behaviors or attitudes they represent have features of hyperactivity in comparison to the whole online population. Not only do they use the Internet more often, but they use its resources more intensively - they more often have social media accounts, buy online more often, make changes to privacy settings more often etc. This is undoubtedly a specific group with a specific attitude towards the online environment. The second important feature that can have a significant impact on the results is the fact that they are professional respondents who get incentives for taking part in the survey. Both these aspects have a particularly significant impact on the results of research carried out using panels¹².

An example of specific preferences of panel users may be IAB Poland 'Privacy in network' survey in which the level of annoyance regarding information banners about cookies was measured. It appeared that in comparison to two other sources of sampling – Run On Site and local Social Networking Service users – the panelists showed an extremely high level of irritation (see: table below).

¹² R.D. Hays, H. Liu, A. Kapteyn, 2015.

ANNOYANCE LEVELS DEPENDING ON SAMPLING**Was the information about cookies you met on the websites annoying for you?**

	Source of interview		
	ROS	Social Media (SNS)	Panel
Definitely no	6%	11%+	1%-
Rather no	15%	18%+	7%-
Neither yes nor no	26%	33%+	20%-
Rather yes	23%	22%-	37%+
Definitely yes	30%+	16%-	35%+
Top 2 boxes:	21%+	29%+	8%-
Bottom 2 boxes:	53%	38%-	73%+
Mean:	3,57	3,15	3,99
Sample size:	1226	979	436

+/- Significant difference on 95% level, tested vs total.

Source: IAB Poland, 'Privacy in network', 2013.

Perspective on the CBA study

Although there is no evidence that the results of the CBA study are skewed due to the selection of the sample, it should be checked whether the phenomenon mentioned above did not occur.

5. Summary

The CBA project is an interesting attempt to evaluate ad experiences on the web and may be the starting point for further research and analysis. The study, however, lacks credibility due to the above mentioned factors regarding: model of exposure, context, relation to the brand, quality of the copy or research sample. As mentioned at the beginning - and what is clearly emphasized in the scientific literature - all experimental studies are only reliable as confirmed by validation. Its lack makes results of the study just an experiment describing fictional ad experience which cannot be a starting point for creating any standards.

Serious doubts are aroused by copies tested in the study. Their low quality and selective choice undermine the credibility of the results. Copies of different quality should be tested - very poor, average and of a high quality. Complex Copy Test including various types of particular ad format should be considered. Major changes are also required in the questionnaire which should measure preferences of users in more objective manner.

Results of the study are debatable as they do not address the local specificity of markets and can lead to serious disturbance of advertising ecosystems. Of course, some results may not raise doubts - in particular, this applies to typically intrusive formats (e.g. unmuted forms). However, even some of the formats can be used in appropriate conditions which are known to be acceptable to the respondent. With such a chosen methodology the final result cannot be a discriminatory ranking. Only general guidelines can be presented on the basis of the research – e.g. which group of ads requires special care and which can be used in a general way.

Having the understanding of the importance of CBA's goals, we would like to offer our help in correcting all the above mentioned areas addressed in this position paper. We are convinced that the research can only be carried on after all the necessary corrections will have been implemented.

References

- Cresswell, J.W. (2014), *Research Design – Qualitative, Quantitative, and Mixed Methods Approaches*; SAGE Publications.
- Friedman, H.H., Wilamowsky, Y., Friedman, L.W. (1981), *A Comparison of Balanced and Unbalanced Rating Scales*; *The Mid-Atlantic Journal of Business*, 19(2).
- Gunter, B. (2000), *Media Research Methods: Measuring Audiences, Reactions and Impact*; SAGE Publications.
- Hays, R.D.; Liu, H.; Kapteyn A. (2015), *Use of Internet panels to conduct surveys*. In *Behavior Research Methods*; Psychonomic Society, Volume 47, Issue 3.
- Judice, T. N; Neuberg, S.L. (1998), *When Interviewers Desire to Confirm Negative Expectations: Self-Fulfilling Prophecies and Inflated Applicant Self-Perceptions*. In *Basic and Applied Social Psychology Journal*; Issue 3.
- Malthora N. K. (2006), *Questionnaire Design and Scale Development*. In *The Handbook of Marketing Research: Uses, Misuses, and Future Advances*; SAGE Publications.
- Schuman, H.; Presser S. (1981), *Questions and answers in attitude surveys: Experiments on question form, wording, and context*; Academic Press.
- van Kleef, E.; van Trijp, H.C.M.; Luning P. (2004), *Consumer research in the early stages of new product development: a critical review of methods and techniques*; Marketing and Consumer Behaviour Group, Department of Social Sciences, Wageningen University.