Prepared for:

Utility Air Regulatory Group

An Assessment of the Robustness of Visual Air Quality Preference Study Results

Prepared By:

Anne E. Smith, Ph.D.
Sabrina Howell

CRA International
1201 F Street, N.W., Suite 700
Washington, DC. 20004

Date: March 30, 2009

TABLE OF CONTENTS

1. BACKGROUND AND SUMMARY OF FINDINGS 1

2. METHODS 3

2.1. SLIDE DEVELOPMENT 3

2.2. PROCESS AND PRESENTATION 7

2.3. PARTICIPANT INFORMATION 9

3. RESULTS 11

3.1. TEST 1: REPLICATION OF ABT (2001) RESULTS 11

3.2. TEST 2: IMPACT OF SHOWING ONLY VAQ DEEMED "ACCEPTABLE" IN ABT SURVEY AND CRA'S REPLICATION OF ABT SURVEY 11

3.3. TEST 3: IMPACT OF SHOWING HIGHER DECIVIEW CONDITIONS 13

3.4. PATTERN OF RESULTS BY RESPONDENT ATTRIBUTES 14

3.4.1. Male/Female 14

3.4.2. Job Position 15

3.4.3. Geographic Location of Residence 17

4. SUMMARY OF OTHER RESPONSES AND OBSERVATIONS OF THE FACILITATOR 18

4.1. RELATIONSHIP BETWEEN SLIDE VAQ, VAQ SCORES AND ACCEPTABILITY RATINGS 18

4.2. ORDERING EFFECT 19

4.3. FOLLOW-UP QUESTIONS 19

4.3.1. Health Concerns 19

4.3.2. Subject Comments 20

5. CONCLUSIONS 24

APPENDIX A: KEY SHOWING ORDER IN WHICH EACH OF ABT'S "SLIDE NUMBERS" WAS PRESENTED IN THE QUESTIONNAIRE 25

APPENDIX B: CRA'S SCRIPT 26

APPENDIX C: DR. SMITH'S EMAIL TO CRA DC STAFF ABOUT CRA'S SURVEY 28

APPENDIX D: RESPONSE BOOKLET USED IN CRA SURVEY 29

APPENDIX E: SUBJECT KEY 38

APPENDIX F: VAQ SCORES BY CRA RESPONDENT 40

APPENDIX G: RATING OF VAQ UNACCEPTABILITY BY CRA RESPONDENT 43

LIST OF FIGURES

Figure 1: Washington DC Vista Used in Abt and CRA Surveys (for a 15.6 deciview level of VAQ) 3

Figure 2: Comparison of Abt Associates (2001) and CRA Results on Percent "Unacceptable" 12

Figure 3: CRA Results on Percent "Unacceptable" by Deciview for "Test 2" (Showing Only Photos Up to 27.1 dv) Compared to Results for "Test 1" (Showing Photos up to 38.3 dv, as in Original Abt Survey) 12

Figure 4: Results for "Test 3" (Showing Photos up to 45 dv) Compared to Tests Showing Less Range in Potential VAQ 14

Figure 5: Male/Female Subject Breakdown of Responses for Tests 1 and 2 15

Figure 6: Job Title Breakdown of Responses for Test 1 16

Figure 7: Job Title Breakdown of Responses for Test 2 16

Figure 8: Office Breakdown of Responses for Tests 1 and 2 17

Figure 9: Average Rankings of Visual Air Quality by Test 19

Figure 10: Evidence of Health Concerns Intertwined in Responses to VAQ Preference Questions 20

Figure 11: Houston Responses to Questions A and C 21

Figure 12: DC Responses to Questions A and C 21

Figure 13: Houston Responses to Question D 23

Figure 14: DC Responses to Question D 23

LIST OF TABLES

Table 1: VAQ Used in Abt Survey and in CRA's Test 1 Version of the Survey 5

Table 2: VAQ Levels of Slides Used in CRA's Test 2 Version of the Survey 6

Table 3: VAQ Levels of Slides Used in CRA Test 3 Version of the Survey 7

Table A-1: Original Key (for Test 1) 25

Table E-1: DC Test 1 Subjects, Interviewed March 2-4, 2009 38

Table E-2: DC Test 2 Subjects, Interviewed March 2-4, 2009 38

Table E-3: DC Test 3 Subjects, Interviewed March 4, 2009 39

Table E-4: Houston Test 1 Subjects, Interviewed February 26, 2009 39

Table E-5: Houston Test 2 Subjects, Interviewed February 26, 2009 39

Table F-1: DC Test 1 Ratings of Visual Air Quality 40

Table F-2: DC Test 2 Ratings of Visual Air Quality 41

Table F-3: DC Test 3 Ratings of Visual Air Quality 41

Table F-4: Houston Test 1 Ratings of Visual Air Quality 42

Table F-5: Houston Test 2 Ratings of Visual Air Quality 42

Table G-1: DC Test 1 Ratings of Acceptability/Unacceptability 43

Table G-2: DC Test 2 Ratings of Acceptability/Unacceptability 44

Table G-3: DC Test 3 Ratings of Acceptability/Unacceptability 44

Table G-4: Houston Test 1 Ratings of Acceptability/Unacceptability 45

Table G-5: Houston Test 2 Ratings of Acceptability/Unacceptability 46

1. BACKGROUND AND SUMMARY OF FINDINGS

The U.S. Environmental Protection Agency (EPA) is considering whether and how to set a secondary National Ambient Air Quality Standard (NAAQS) for particulate matter (PM) that would address impacts to visual air quality (VAQ), or "visibility." The emphasis of this deliberation is presently on protection of "urban visibility," which is intended to refer to the aesthetic impacts that people experience in their day-to-day lives in the area where they live and work. A type of survey that EPA calls a "VAQ preference study" has taken a central role in EPA's deliberations. EPA appears to view VAQ preference studies as more reliable for standard-setting than studies that attempt to estimate the public's willingness to pay to improve VAQ. (EPA refers to these latter studies as "visibility valuation studies," which include both stated and revealed preference methodologies.) Visibility valuation studies do, indeed, present many technical difficulties and credibility concerns, but they have at least been subjected to long and intense technical scrutiny. VAQ preference studies, in contrast, have not been subjected to the same degree of scrutiny. More thorough research should be conducted on the robustness and technical credibility of results from VAQ preference studies if they are to be used in national standard-setting. This report describes one such research effort conducted by CRA International (CRA).

We reviewed the four existing VAQ preference studies cited in EPA's PM NAAQS review documents, and hypothesized that their results may be driven by the *implicit suggestion* in the questionnaire that some levels of the VAQ in the survey may be unacceptable, and not by individuals' internally-held opinions about what absolute level of visibility degradation is unacceptable. If our hypothesis is correct, research that performs one or more controlled variations on the survey design, in which only the range of VAQ shown to the participants is changed, should find that answers to the VAQ preference survey will change in a predictable manner across those variants. If this result is found, then the results of VAQ preference studies cannot be considered a robust basis for setting a level for a secondary NAAQS.

During February and March 2009, CRA conducted research providing a preliminary exploration of this hypothesis, surveying 64 individuals randomly assigned to three variants of the same VAQ preference study. We based the instrument design on the VAQ preference study pilot conducted by Abt Associates (Abt) in 2001 for a Washington DC vista.[1] Twenty-six of the individuals in CRA's study were presented a questionnaire and photographs that were nearly identical to those used by Abt. This sample establishes a baseline for comparing results from two variants of the design, and also to establish whether the CRA survey was a reasonable replication of the Abt pilot survey. The other 38 individuals interviewed were subjected to one of two variants in which the range of VAQ shown in the photographs was either narrower or wider than Abt used. We found that the pattern of responses about "unaccept-

[1] Abt Associates, Incorporated (2001) *Assessing Public Opinions on Visibility Impairment Due to Air Pollution: Summary Report*, work funded through the U. S. Environmental Protection Agency, Contract No. 68-D-98-001, Work Assignment 3-53 (January).

ability" changed in a pronounced manner when we changed only the range of VAQ presented. We conclude that the results of VAQ preference studies are *not* robust and do not reflect an enduring view on the "unacceptability" of different levels of VAQ degradation.

Although it is only preliminary exploration of this particular concern with the VAQ preference methodology, CRA's research suggests that a majority of the sampled people will *always* identify one or more of the VAQ levels shown to them as "unacceptable." It appears that the cutoff of "median unacceptability" is determined more by the range of VAQ shown in the questionnaire than by any absolute notion held by the public of an unacceptable degree of VAQ degradation. This is a serious flaw in the methodology that must be addressed in a thorough and scientific manner before VAQ preference studies can be deemed a more valid basis for setting visibility standards than the problematic visibility valuation studies.

Results from our survey also reinforce concerns identified in the Abt study that respondents to VAQ preference studies intertwine health risk concerns with aesthetic concerns, even when instructed to consider only the aesthetic concerns. This has long been recognized as one of the difficulties in the interpretation of results from visibility valuation studies, but VAQ preference studies are not immune to this methodological problem either.

The rest of this report documents the methods that CRA employed in its research on this matter, and provides details of the results. We recognize that this study should be categorized as preliminary research; however, its findings are so pronounced that the question of robustness that it has identified cannot be dismissed without more in-depth study, and methodological changes are found that resolve it in a satisfactory manner.

2. METHODS

2.1. SLIDE DEVELOPMENT

Abt Associates used a set of 35 mm slides of a vista of Washington DC viewed from across the Potomac River near Arlington Cemetery. The vista included the Mall in downtown Washington, DC, and landmarks including the Lincoln Memorial, Washington Monument, Capitol Building, Union Station, and Library of Congress. Figure 1 presents the vista that Abt used (and which CRA also used). Using electronic techniques, this single picture was altered to display varying levels of visibility that were then combined into a set of 25 photographs that Abt used in its questionnaire.[2]

Figure 1: Washington DC Vista Used in Abt and CRA Surveys (for a 15.6 deciview level of VAQ)

According to the Abt Associates report, "the EPA recommended that the levels of $PM_{2.5}$ represented in the slides range from estimated natural conditions (about 2.3 $\mu g/m^3$) to high haze conditions (65 $\mu g/m^3$) in order to simulate a broad range of pollutant concentrations."[3]

[2] Abt's slides were produced by Air Resource Specialists (ARS) using a modeling technique that is also possible to perform in a more simplified manner with ARS's "WinHaze" software.

[3] Abt Associates (2001), p. 4.

In order to convert these $PM_{2.5}$ concentrations into specific levels of VAQ for the slides used in the survey, the $PM_{2.5}$ species were assumed to be present in constant proportions, and a single relative humidity was also assumed.[4] In this way, Abt could state that the differences in visual air quality in the different slides were solely due to changes in PM mass concentrations. While this is a true statement, we note that none of the individual photographs can be said to reflect a unique specific $PM_{2.5}$ concentration. Even for a fixed pollutant mix and uniform spatial distribution of the pollutants, each individual $PM_{2.5}$ level can be consistent with a range of VAQ outcomes due to normal amounts of fluctuations in relative humidity levels. For this reason, the best way to describe the individual slides used is in terms of one of the three VAQ metrics.[5] In the rest of this report, therefore, we will only describe the slides used in each of the survey variants in terms of their VAQ, using the deciview (dv) metric.

Table 1 (which is a direct copy of Exhibit 2 from the Abt report) lists the set of 20 slides used in the Abt Associates study and their associated visibility metrics. Though there are only 20, they number 1 through 25 because Abt Associates originally intended to use slides showing 25 VAQ levels. However, before conducting their actual survey, they eliminated five slides at the highest end of the dv range, because they found that people could not tell the difference between adjacent slides with high dvs. In their place, Abt inserted "repeats" of five of the remaining 20 slides to create their final "test set" of 25. The slides that were repeated were numbers 5, 13, 22, 23 and 25. The final test set was randomly ordered, as documented in Appendix A of the Abt report (reproduced in Appendix A of this report).

CRA used ARS's WinHaze 2.9.0 software to recreate electronic file versions of each of the 20 deciview levels listed in Table 1 using the same Washington DC vista that Abt used (Figure 1). These were ordered exactly as in Abt's survey, including the five duplicate slides. This set of 25 photographs was used for CRA's "Test 1," a nearly-exact replica of the VAQ preference questionnaire that Abt used in its study. In terms of the photographs of VAQ used, the differences were that CRA used the simpler WinHaze software and projected the results onto a screen from the laptop, rather than using high-quality 35 mm slides. One might have expected that the differences in the quality of the photographic displays used in our survey could interfere with the replication, but the results presented in Section 3 of this report should dispel any such concern.

---

[4] Based on data from the Washington DC IMPROVE monitoring site between 1988 and 1999, they assumed the pollutant mix to be: sulfate = 50%; nitrate = 10%; organic carbon = 25%; elemental carbon = 10%; fine soil = 5%. Abt does not report what relative humidity was assumed, but it appears (from the deciviews reported to go with each $PM_{2.5}$ level) that it was about 73%.

[5] Visual air quality is often described by one of three mathematically-related metrics. These metrics are visual range, expressed in kilometers or miles; the light extinction coefficient (Bext), expressed in inverse megameters (Mm^{-1}); and the deciview (dv), which expresses changes in visual air quality that are incremental in terms of perception across the range of conditions. A lower dv indicates higher VAQ, which is synonymous with "greater visual range," "less light extinction," and "better visibility."

Table 1: VAQ Used in Abt Survey and in CRA's Test 1 Version of the Survey

Slide Number	$PM_{2.5}$ Concentration ($\mu g/m^3$)	Visual Range (km)	Deciview (dv)	Bext (Mm^{-1})
Slide 1	65	8.5	38.3	460
Slide 3	55	10.0	36.7	391
Slide 5	50	11.0	35.7	357
Slide 7	45	12.1	34.7	323
Slide 9	40	13.6	33.6	288
Slide 11	35	15.4	32.3	254
Slide 12	32.5	16.5	31.6	237
Slide 13	30	17.8	30.9	219
Slide 14	27.5	19.3	30.1	202
Slide 15	25	21	29.2	185
Slide 16	22.5	23	28.2	168
Slide 17	20	26	27.1	151
Slide 18	17.5	29	25.9	133
Slide 19	15	34	24.5	116
Slide 20	12.5	39	22.9	99
Slide 21	10	48	21.0	82
Slide 22	7.5	60	18.7	65
Slide 23	5	82	15.6	48
Slide 24	2.5	129	11.1	30
Slide 25	2.32	163	8.8	24

In terms of the questionnaire used, we replicated only the portion of the Abt survey that contained questions consistent with a VAQ preference study. We did not ask the Abt survey questions that came later in its script regarding length of time that a particular VAQ level would have to occur in order to be unacceptable, or about willingness to pay. These were not part of any other VAQ preference studies, and also were not of interest for exploring our hypothesis concerning the VAQ preference study method.

As noted in Section 1, CRA's goal was not simply to explore whether it could reproduce the Abt survey results. Rather, CRA's goal was to explore the sensitivity of the VAQ preference survey results to controlled variations in the range of VAQ shown to the sampled individuals. Thus, after reconstructing the set of 25 photographs used by Abt, two additional sets of photographs were created. CRA's "Test 2," showed only VAQ levels that had been found to be "acceptable" by a majority of the respondents in the Abt study. That is, Test 2 was performed with the subset of 12 slides from Test 1 that had less than 27.1 dv, which were (from Table 1) Slides 17 through 25, plus the three repeat slides that were in this narrower dv range (which were the repeats of Slides 22, 23 and 25). The Test 2 slides were kept in the same order as the Test 1 slides, by simply omitting showing any of the higher dv slides.

Table 2: VAQ Levels of Slides Used in CRA's Test 2 Version of the Survey

(There are three "repeat" slides among the 12 slides of the Abt survey with less than 27.1 dv, and they have been noted in the table below as the original slide number followed by the suffix .1)

Abt Slide Number	Deciview
17	27.1
18	25.9
19	24.5
20	22.9
21	21
22	18.7
22.1	18.7
23	15.6
23.1	15.6
24	11.1
25	8.8
25.1	8.8

We also created a third set of slides that presented a wider range of visibility than in the original Abt survey. In the original study, the highest dv shown was 38.3 dv. For CRA's "Test 3" variant of the survey, two photographs with higher dvs were added. These two new slides presented VAQ conditions of 42 dv and 45 dv, while three of the original dv slides were removed from within the range, so that Test 3 still showed the lowest dv level as in the original set (*i.e.*, 8.8 dv), but showed respondents a wider range of dvs without increasing the total number of slides presented. The omitted slides were Slides 19, 23, and 24 (See Table 1), which had 24.5 dv, 15.6 dv, and 11.1 dv, respectively. In the interests of time,

the repeat slides were not included in Test 3 (they had not been revealing any concerns with inconsistencies in responses). Table 3 shows the specific VAQ levels used in CRA Test 3. Also, the slides were shown in a different random order than in the original survey. The order used in Test 3 is identified in the third column of Table 3.

Table 3: VAQ Levels of Slides Used in CRA Test 3 Version of the Survey

Abt Slide Number	Deciview	Order in which slide presented during Test 3
1.1 (CRA's New)	45	4
1.2 (CRA's New)	42	15
1	38.3	11
3	36.7	18
5	35.7	2
7	34.7	10
9	33.6	9
11	32.3	7
12	31.6	14
13	30.9	17
14	30.1	1
15	29.2	12
16	28.2	16
17	27.1	8
18	25.9	6
20	22.9	19
21	21	3
22	18.7	13
25	8.8	5

2.2. PROCESS AND PRESENTATION

Abt conducted its survey in a group format, with all the respondents answering the survey questions at the same time. CRA took a different approach, which was to conduct the survey in a sequence of one-on-one interviews. This allowed us to glean individual comments about the test during follow-up questions. We relied solely on CRA staff for our samples.

A single facilitator (Ms. Sabrina Howell) performed all of the interviews, following a script which is provided in Appendix B. The facilitator was an employee from CRA's Houston office. She had no previous experience with visibility issues, nor with air quality policy or research generally. Being from the Houston office, she was not known by most of the Washington DC CRA staff to whom she administered the survey. However, no attempt was made to limit the interviews to only people who did not know her. Dr. Smith identified the general approach, including specifying how to construct the variants called Test 2 and Test 3. She did not, however, suggest to Ms. Howell whom she should ask to participate, and did not attend any of the interviews or talk to any participant about this research prior to their interview. The only communication from Dr. Smith to CRA staff about this survey was in an email that she sent

on the day of the majority of the DC office interviews to explain that they might be approached by Ms. Howell. A copy of that email is provided in Appendix C.

An initial pilot of Tests 1 and 2 was conducted on February 20, 2009 with eight people who work in CRA's DC office. The first four interviews used Test 1, followed by four interviews using Test 2. Results from these eight interviews were reviewed to determine whether there was any merit in continuing the survey with a larger sample. The initial pilot showed a pronounced difference in the median point of unacceptability, even for samples of only four individuals each. We therefore decided to continue with interviews. No other CRA staff were informed about the results of the pilot, and none of the eight individuals who had been interviewed were told about the purpose of the survey, or that they had been given different versions of the survey, until after all of the interviews were completed.

A second round of interviews using Test 1 and Test 2 was conducted on February 26, 2009 in CRA's Houston office, with CRA staff who work in that office. Although the survey was conducted with Houston residents, they were still shown the same Washington DC photograph. We performed the survey with Houston residents because we wanted to be able to explore whether the hypothesized dependence of responses to the range of VAQ levels in the survey would exist regardless of whether the sampled parties were personally familiar with typical Washington DC VAQ levels. In the Houston office, ten subjects completed Test 1, and ten subjects completed Test 2, randomly assigned. We also started to ask some follow-up questions, to be able to explore the thinking underlying the responses.

The final round of interviews was conducted in the DC office on March 2-4, 2009. Given that few people in the office knew Ms. Howell, we decided on March 3 that Dr. Smith should send out the introductory email in Appendix C. We also introduced Test 3 on March 3. A total of 36 people were interviewed in the DC office on March 2-4, randomly assigned to Test 1, 2 or 3. Follow-up questions were given after these interviews as well.

Therefore a total of 64 people were surveyed, with a total of 44 responses from CRA staff in the DC office, and 20 responses from CRA staff in the Houston office. Similarly, we had a total of 26 responses to Tests 1 and 2, respectively, and 12 responses to Test 3.

Subjects marked their responses in the same booklet that Abt Associates used, which CRA had photocopied from the Abt report (see Appendix D). They were each guided by the facilitator in the same way, according to the script (see Appendix B). The facilitator did answer clarifying questions when asked, but stayed with the script except for such occasional interruptions, and did not directly encourage such questions. For the first eight DC respondents and the Houston respondents, the visual presentation of the VAQ conditions was done at a large desktop computer monitor, with the subject seated directly in front of the monitor, and the facilitator controlling the slide progression. The rooms used were windowless, but did have ambient lighting. The second set of DC respondents were shown the same slides but projected onto a screen in a darkened, windowless conference room. Results indicate that the manner of presentation of the VAQ conditions did not alter the responses (see Section 3).

The general flow of the interview was that subjects first viewed Slide 23, which has a VAQ of 15.6 dv. (Figure 1 is a copy of Slide 23.) They were then asked a few generic questions about how and when they see a similar degree of visual air quality in their daily life and their beliefs about visibility, following the Abt script. The facilitator then showed them the full set of slides one by one, and had respondents score the VAQ in each one on a scale from 1 (very poor) to 7 (very good). They were not allowed to go back and view slides again in order to reevaluate their earlier judgment. The facilitator then showed them the same set of slides once more, and this time they were asked to rate each slide either as "acceptable" or "unacceptable." Finally subjects were asked a few follow-up questions about what they were thinking during the process and whether or not they had had thoughts about health concerns. Follow-up responses were the only responses given orally, and recorded by the facilitator.

2.3. PARTICIPANT INFORMATION

As described above, all the subjects for our tests were CRA employees in the firm's Washington, DC and Houston offices. They were solicited at random by the facilitator, depending on whom the facilitator would first encounter in the office who was not obviously engaged in a meeting or on the telephone. Thus, solicitation of people to be interviewed was not randomized in a scientific *a priori* manner, but neither was it systematic in any way. If a staff member was not in the office on any of the days when the facilitator was conducting interviews, they would not have been solicited. Also, some staff would decline to take the interview, usually because they were too busy. No attempt was made to follow up with them, but the intent of the email from Dr. Smith was to minimize the number of people who declined. In the end, the 64 people who were interviewed included both research and non-research staff, from entry-level to senior management, and working in various practice areas. While they do not reflect a scientifically-randomized sample of CRA employees, they do reflect a broad spectrum of the staff in the DC and Houston offices. Appendix E provides detail on the sex and job category of all individuals interviewed, broken out by test date and test variant taken.

In the broader context, the CRA sample is certainly not a demographically representative sample of the populations of either Houston or Washington DC. First, all respondents (obviously) were employed. Second, the sample is almost certainly more highly educated and has a higher range of household income levels than in the general population. However, the purpose of this research was to explore the *robustness* of the VAQ acceptability results to changes in the survey instrument, and it was not to produce a new estimate of a VAQ unacceptability cutoff for use in any policy decision. The question of robustness can be assessed without using samples that are representative of the broader population as long as the subgroups subjected to the different versions of the survey are drawn from the same pool of individuals, and do not differ markedly in their respective makeup. We accomplished that objective by randomly assigning people who agreed to be interviewed to the test version that they were then given. Evidence of the degree to which each test was given to a similar mix of the CRA population can be found in the detailed tables of Appendix E.

The VAQ scores and acceptability responses of each individual are provided in Appendices F and G, respectively. Sections 3 and 4 of this report provide various summaries of these raw data, and of the individuals' responses to other questions asked of them.

3. RESULTS

3.1. TEST 1: REPLICATION OF ABT (2001) RESULTS

CRA's Test 1 tried to recreate the Abt Associates pilot survey. This was intended to establish a baseline, as well as to determine whether our quick survey would reproduce the results found by Abt Associates. Figure 2 compares results from CRA's Test 1 with the Abt results. In Figure 2, each point corresponds to a slide with a certain VAQ, stated in dv (identified on the x-axis). The y-axis shows what percentage of subjects who took this version of the survey determined that the each VAQ was "unacceptable." In both the Abt and CRA Test 1 survey, the VAQ that 50% of respondents deemed to be "unacceptable" is between 27 and 30 dv. This represents a remarkable degree of replication, given the differences in the samples, and the relatively small sample size of the Abt pilot (nine people). Most remarkable of all is that both surveys produced the same pronounced dip in the unacceptability rating at 28.2 dv (Abt's Slide 16), which causes the rather wide range of median unacceptability. Abt hypothesized that this might be a result of slide ordering. Given that we used the same slide ordering, this suggests that Abt's hypothesis is correct.[6]

Figure 2 indicates that differences of venue, sample selection methods, and visual presentation of the photographs with varying VAQ levels have not undermined CRA's ability to perform a VAQ preference test comparable to Abt's. This frees us to use our own survey to test the robustness of results to instrument design. The most interesting findings of CRA's study follow from the comparisons of Test 1 to CRA's Test 2 and Test 3.

3.2. TEST 2: IMPACT OF SHOWING ONLY VAQ DEEMED "ACCEPTABLE" IN ABT SURVEY AND CRA'S REPLICATION OF ABT SURVEY

As can be seen in Figure 2, photographs with VAQ less than 27.1 dv were deemed by a majority of respondents in the Abt survey to be "acceptable." (That is, less than 50% deemed them "unacceptable.") In Test 2, CRA asked a different set of 26 individuals to go though the same exercise, but showing them only photographs with 27.1 dv or less. If the original survey results had revealed an absolute notion of VAQ unacceptability, then the second of CRA's sample groups should not have assigned more than 50% unacceptability to any of the photographs they were shown. Only the worst one or two of the VAQ levels shown to them should have been even in a mid-range of unacceptability ratings. However, Figure 2 shows that the Test 2 sample *determined that there was a point of unacceptable VAQ well within the range of VAQs that it was shown.*

---

[6] See Abt (2001), p. 21. Slide 16 (with mid-range VAQ of 28.2 dv) is the second slide shown in the survey, directly after Slide 1, which has the highest dv rating of 38.3 dv. One can hypothesize that it "looked" relatively much better to survey respondents if they are asked to rank what they know to be the worst-VAQ in the entire questionnaire just prior. (They "know" that slide 1 is the worst-case VAQ when they answer the acceptability questions because the preceding scoring on a scale of 1 to 7 in the questionnaire has informed them of this fact.)

Figure 2: Comparison of Abt Associates (2001) and CRA Results on Percent "Unacceptable" (Abt results from Abt's Exhibit 11, averaging scores for slides that were rated twice)

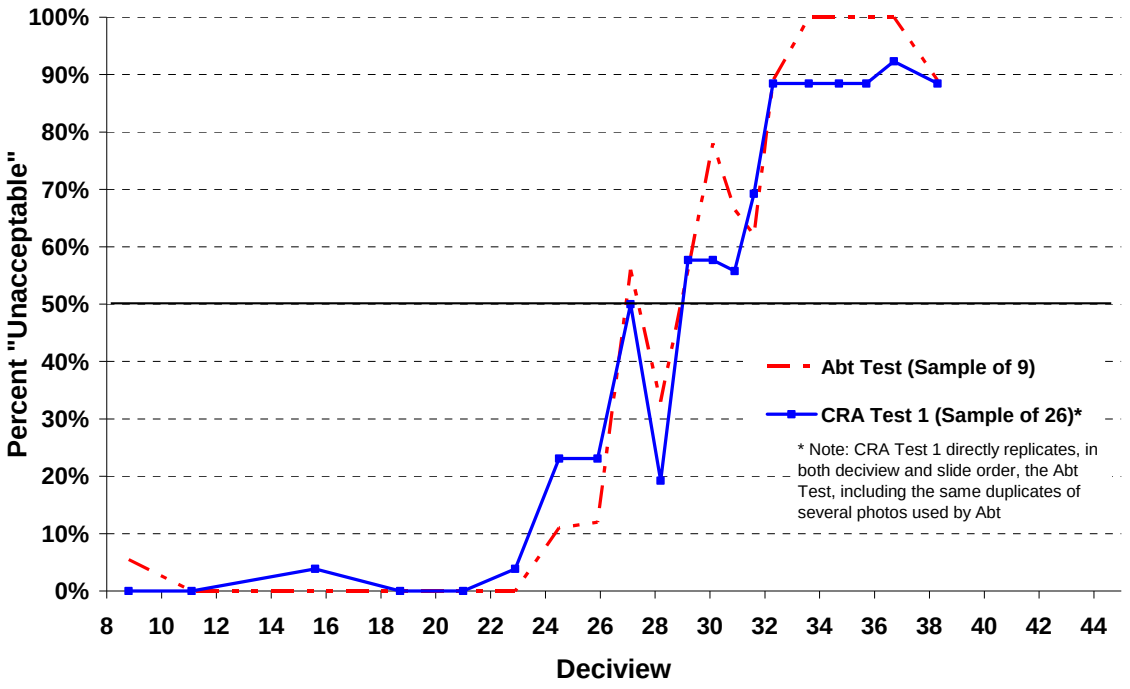
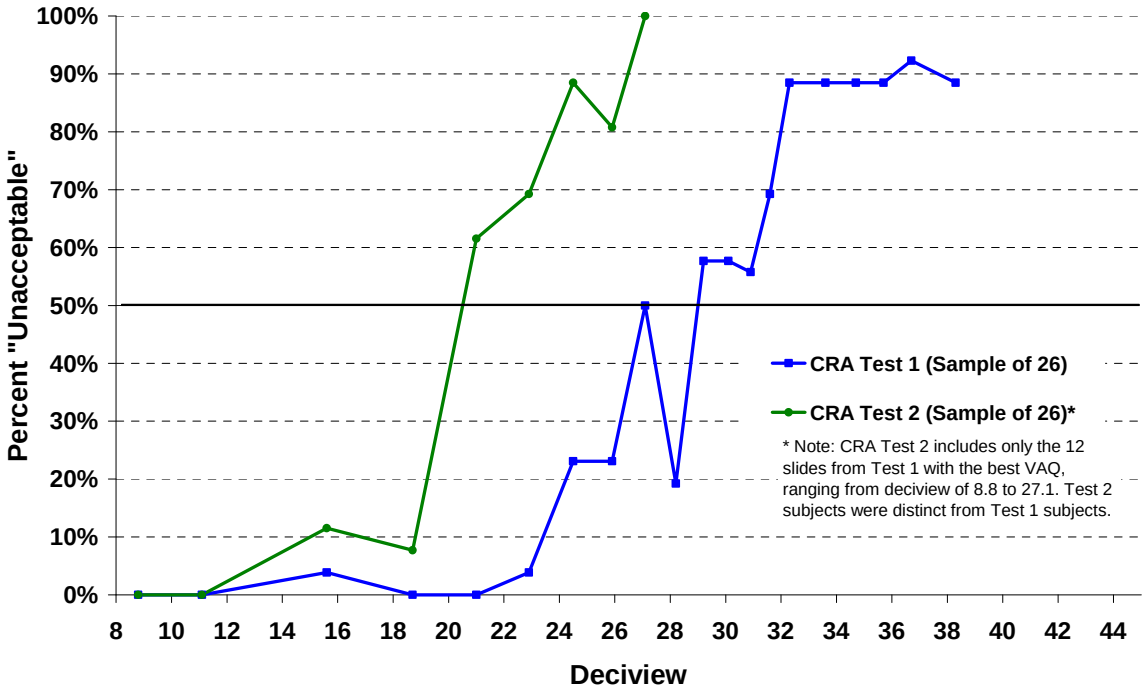


Figure 3: CRA Results on Percent "Unacceptable" by Deciview for "Test 2" (Showing Only Photos Up to 27.1 dv) Compared to Results for "Test 1" (Showing Photos up to 38.3 dv, as in Original Abt Survey)



The most remarkable result apparent in Figure 3 is that the five highest-dv conditions shown in Test 2 were deemed "unacceptable" by more than 50% of the Test 2 sample, even though they were rated "unacceptable" by as few as 0% of the respondents in Test 1. There was no difference in the population sampled for Tests 1 and 2, and the only difference in the questionnaire that they faced was that the Test 1 respondents were also shown a number of photographs depicting other, worse VAQ conditions than were shown to Test 2 respondents. The VAQ for a 50% unacceptability rating fell from the 27 to 30 dv range found by both Abt and CRA's replication of Abt to less than 21 dv in the Test 2 variant. There is approximately a 100% difference in the median unacceptable dv levels inferred from Tests 1 and 2, respectively.

These results strongly support CRA's hypothesis that respondents to VAQ preference studies may be induced to feel that at least one of the photographs they are being shown "must be" unacceptable. Whether this is the underlying explanation or not, the comparison of results from Test 1 and Test 2 shown in Figure 2 provides striking evidence that results from VAQ preference studies are not reliably revealing an absolute standard for visual air quality held by the public.

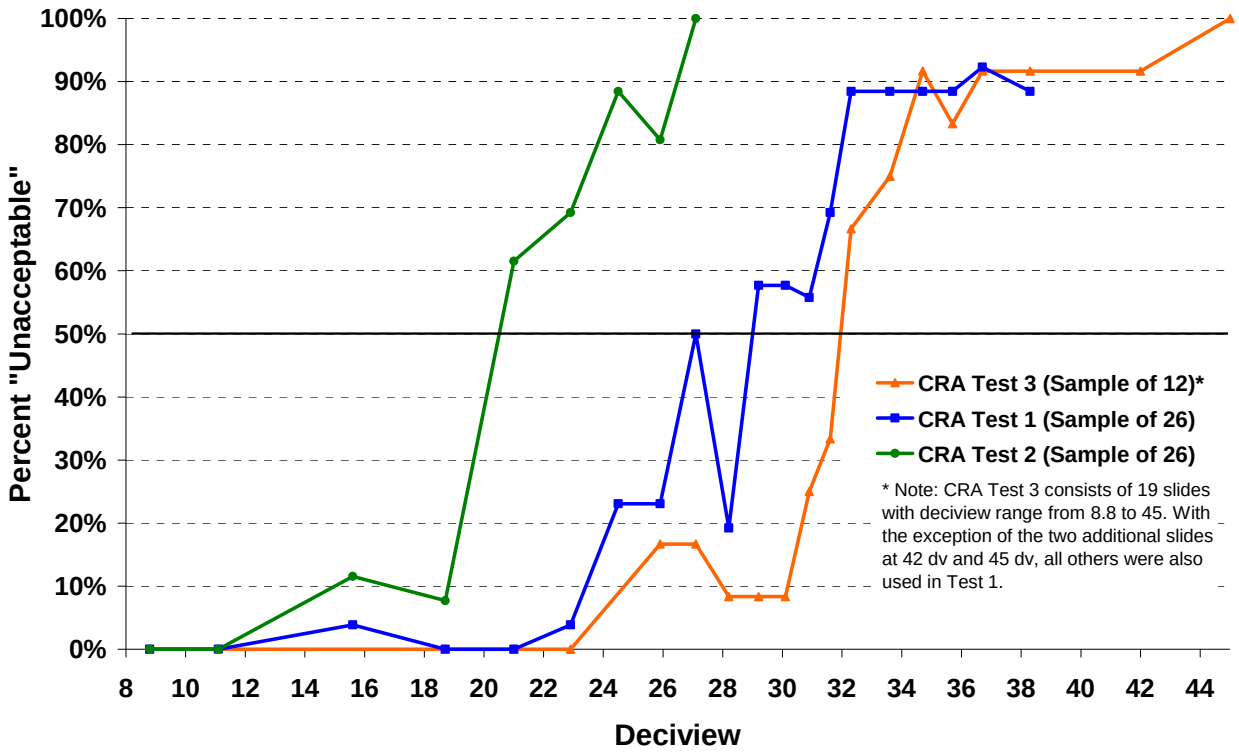
3.3. TEST 3: IMPACT OF SHOWING HIGHER DECIVIEW CONDITIONS

If the interpretation of the comparison of Test 1 and Test 2 is that the average "unacceptability" rating of a VAQ level is determined by the range of VAQs shown to those being interviewed, then one should expect to observe this same effect also appear in the opposite direction. That is, if a *wider* range of VAQ were to be shown, one should see that levels of VAQ deemed "unacceptable" in the original survey will become "acceptable" in a survey that shows photographs depicting more visibility degradation. To check if this is the case, we created Test 3, which included two photographs with higher dvs (*i.e.,* greater VAQ degradation) than the highest dv shown in the original Abt survey. Test 3 was administered to a sample of 12 individuals in the Washington DC office.[7]

Figure 4 compares the results for Test 3 to those for both Test 1 and Test 2. These results also are consistent with our hypothesis that the unacceptability ratings from a VAQ preference study are almost entirely contingent on the range of visibility levels included in the survey instrument. The unacceptability rating of all VAQ levels up to 34.7 dv are consistently lower in Test 3 than in Test 1. For the VAQ deemed about 50% unacceptable under Test 1 (*i.e.,* 27 to 30 dv), the VAQ was considered unacceptable by only about 10-15% of the Test 3 respondents. The 50% unacceptable level rose from 27 to 30 dv in Test 1 up to about 33 dv in Test 3.

[7] Compared to a maximum of 38.3 dv in the Abt survey, photos in CRA's Test 3 also included photos of 42 dv and 45 dv. These VAQ conditions can occur even under attainment of the current $PM_{2.5}$ standard on non-rainy but high relative humidity days.

Figure 4: Results for "Test 3" (Showing Photos up to 45 dv) Compared to Tests Showing Less Range in Potential VAQ



3.4. Pattern of Results by Respondent Attributes

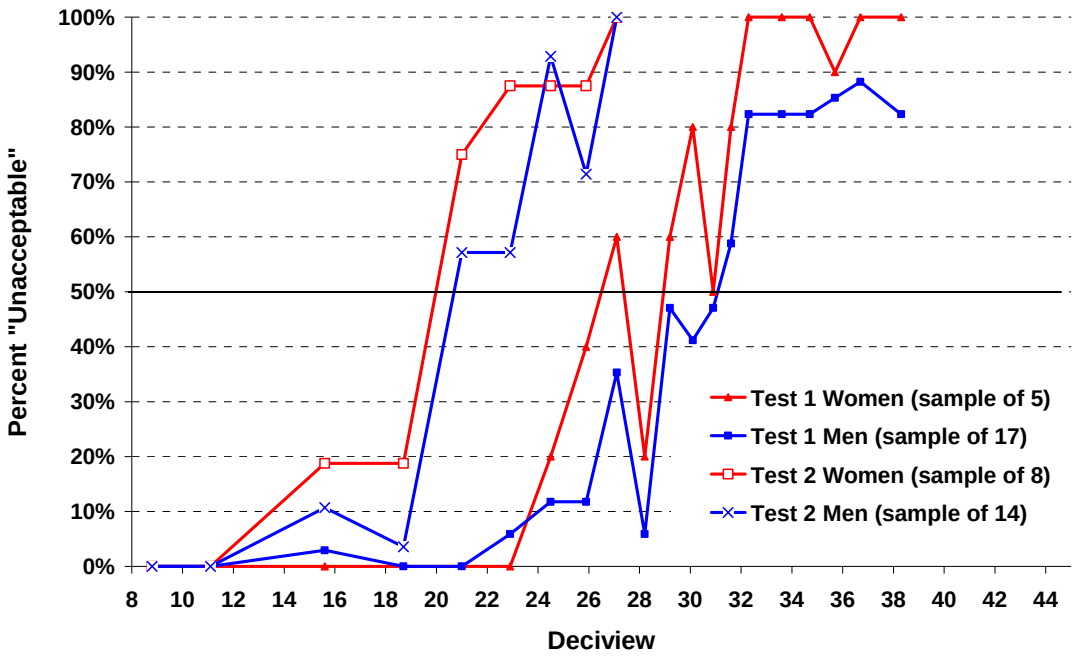
The following demographic breakdowns do not include the pilot group because demographic information in the first round of interviews was not maintained in a manner that could link it back to the individual responses. The pilot group, however, included two women and six men, and all the subjects were CRA research staff. This set of eight responses produced the same pattern in which VAQ levels in Test 2 had higher unacceptability ratings than in Test 1, but we cannot categorize these results by the various demographic breakdowns.

3.4.1. Male/Female

To analyze whether there were any major demographic differences in the sensitivity of unacceptability ratings to test variant, we broke the respondents into several demographic categories. The first, male and female, is shown in Figure 5. It seems that women on average may have a greater inclination to identify each VAQ level as “unacceptable.” However, the pattern of responsiveness to VAQ range shown is equally pronounced for both men and women. This is apparent even though far more men than women were surveyed.

Also notable in Figure 5 is that the anomalous decrease in the acceptability rating of the slide with 28.2 dv appears in both the male and female sample. Further, the worst case VAQ in Test 2 was deemed unacceptable by 100% of both men and women, even though it was 35% to 60% unacceptable to men and women, respectively, under Test 1.

Figure 5: Male/Female Subject Breakdown of Responses for Tests 1 and 2



3.4.2. Job Position

Figures 6 and 7 break down respondents by their job positions for Test 1 and Test 2, respectively. CRA employees are categorized into three general categories: non-research staff (which includes staff in office services, information technology services, and executive assistants); entry-level research staff (which includes Analysts and Associates); and senior research staff (which includes Senior Associates, Senior Consultants, Principals, and Vice Presidents). These categories do not necessarily provide a good proxy for either household income or education level, but consideration of how our patterns of results vary across these very general subgroups of our samples is instructive. The results in Figures 6 and 7 show that the pattern of responsiveness to test variant is apparent in every single group. The findings presented in Figures 2 through 4 above do not appear to be driven by any single subgroup, nor does any single subgroup appear to respond differently from the average. Most remarkably, 100% of all four job categories deemed the 27.1 dv case to be unacceptable under Test 2, even though no more than 50% of any of the four subgroups considered it unacceptable under Test 1.

Figure 6: Job Title Breakdown of Responses for Test 1

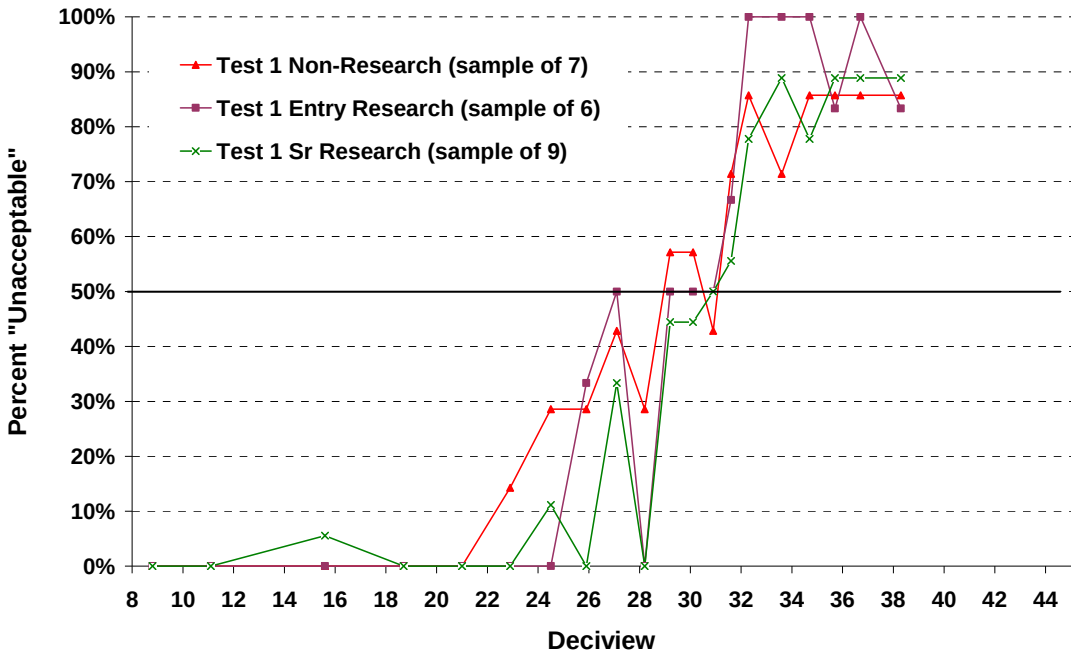
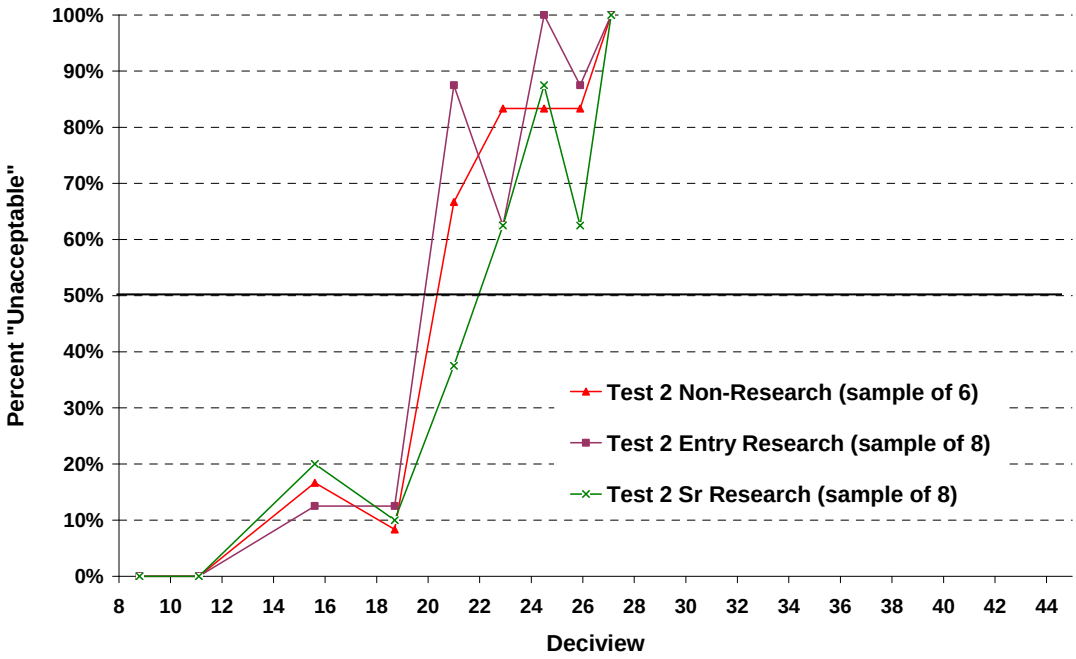


Figure 7: Job Title Breakdown of Responses for Test 2

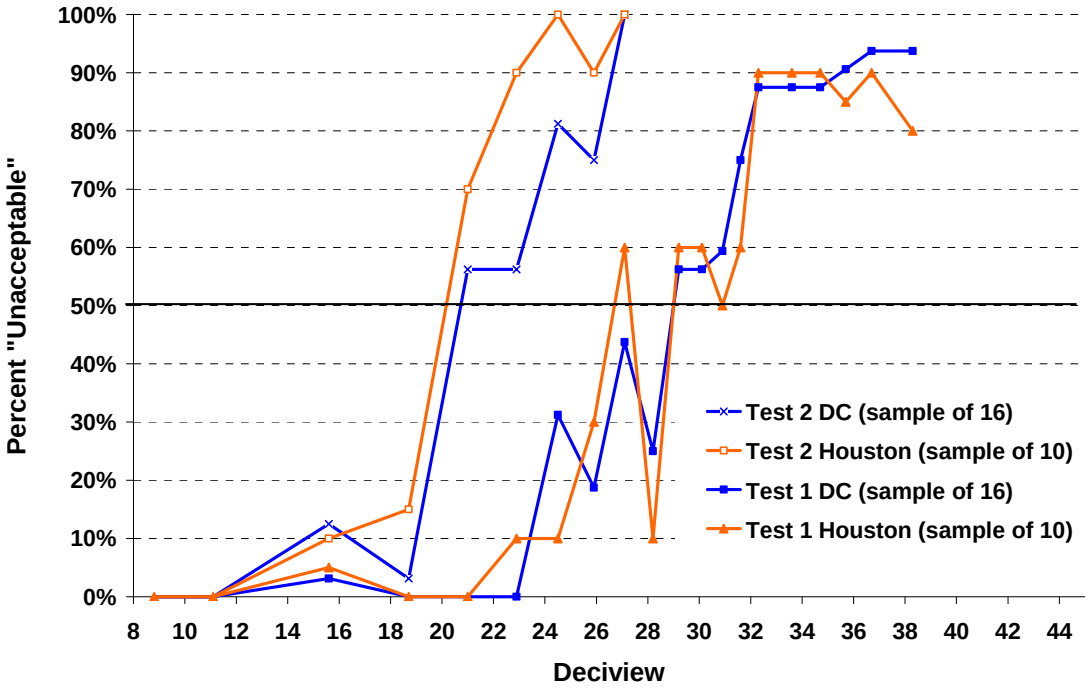


3.4.3. Geographic Location of Residence

We also did not find a notable difference between the Houston office and the DC office responses, as shown in Figure 8. Regardless of location of domicile, 100% of respondents deemed 27.1 dv to be unacceptable when it was the worst-case VAQ shown (Test 2), while it was deemed roughly 50% unacceptable when it was just the mid-range of VAQ shown (Test 1).

(Also, note again that both the Houston respondents and the DC respondents reported the anomalous dip in unacceptability for the 28.2 dv photograph. This again indicates the importance of slide ordering in responses.)

Figure 8: Office Breakdown of Responses for Tests 1 and 2



There were presentation differences between the offices as well. The Houston sample was interviewed at a desk, looking at the slides on a large computer monitor. The DC sample of March 2-4 was surveyed in a darkened conference room, looking at the slides projected onto a large screen. The lack of any obvious difference between the results for either office group shown in Figure 8 indicates that the presentation of the photographs also does not seem to alter our fundamental finding that VAQ preference studies results are not producing reliable insights about a VAQ unacceptability cutoff point.

4. SUMMARY OF OTHER RESPONSES AND OBSERVATIONS OF THE FACILITATOR

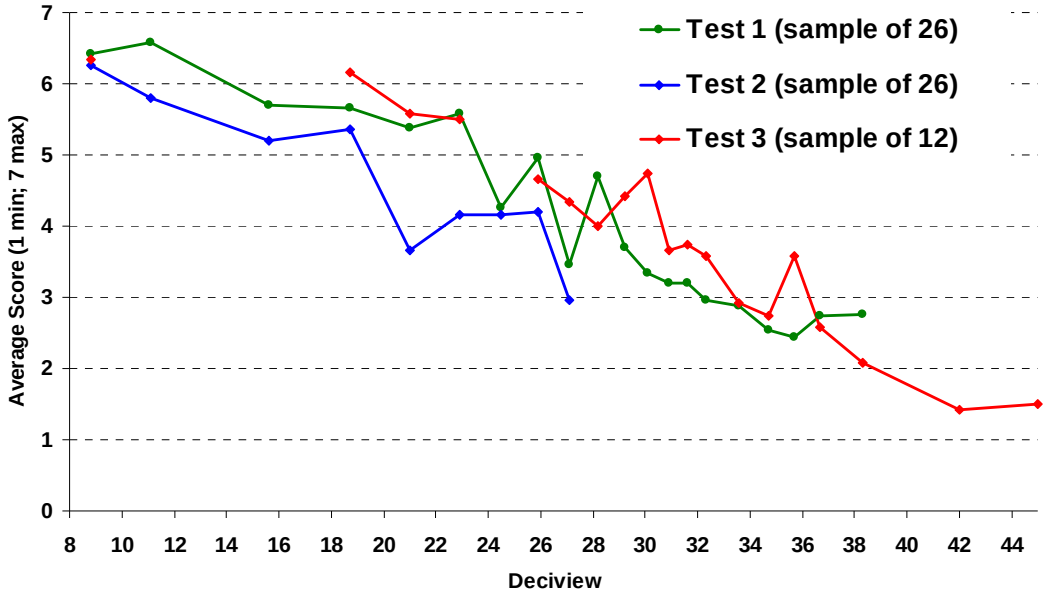
4.1. RELATIONSHIP BETWEEN SLIDE VAQ, VAQ SCORES AND ACCEPTABILITY RATINGS

The subjects revealed sound abilities to perceive variations in visibility conditions. At lower dv levels, they scored the VAQ at or near a 7 ("very good") while increasingly at higher dv levels, respondents consistently scored the view lower and lower. Figure 9 graphs the average scores for each dv level. The y-axis is average of the VAQ scores across all subjects who took each test version. The average scores can only range between 1 ("very poor" VAQ) and 7 ("very good" VAQ).

Regardless of test version, Figure 9 shows that respondents are capable of discerning differences in visibility conditions when shown photographs of the same vista with varying VAQ, consistent with what other VAQ preference studies have reported. We did not question that this type of result would be found. However, what is interesting is that the act of showing a narrower range of VAQ level than experienced in real-life in Washington DC did not induce respondents to assign scores as low as 1 to the worst photograph shown. For example, Test 2 respondents' lowest average score was about 3 for the 27.1 dv slide, even though this was the worst-case VAQ that they were shown. Given the vastly higher percent unacceptability rating that they assigned that VAQ level than did the Test 1 and Test 3 respondents, one might have expected to have seen the Test 2 group to have "stretched" their mental scoring scale too, so that the worst case photograph would have had a minimum score, closer to 1 or 2. This did not happen. Each of the 3 groups seemed, on average, to understand whether the range of VAQ being shown to them during the interview was reflective of the full range of VAQ that they see in their daily lives, and to preserve the lowest scores only for conditions that they know from experience to be truly worst-case in the real world.

Nevertheless, there are some differences in the scores by test variant. Although all three test groups gave a very similar average ranking to the photograph with 8.8.dv, their rankings do not all slope down at the same rate as the dv level increases. As a result, we find a very similar average score associated with the various percent unacceptability responses. For example, the median cutoffs of unacceptability for Tests 1, 2 and 3 were about 20 dv, 27 dv and 32 dv, respectively, but the average VAQ scores at these respective dv levels are about 3.5 in all three cases. It is notable that 3.5 is exactly half of 7, or a "perfect score." This suggests that perhaps people, when attempting to decide what "unacceptable" means, are deciding to define it as anything that they perceive as below the average within the range of VAQ they are currently experiencing. This possibility should be studied in further research.

Figure 9: Average Rankings of Visual Air Quality by Test



4.2. ORDERING EFFECT

In Figure 2 it is clear that not only did the CRA Test 1 match the general results of Abt Associates' survey, *it also matched the outliers.* We believe that the ordering of the photos produces the outliers. The outlier occurred at Slide 16, with 28.2 dv. For Test 1, this slide was shown second. The first slide was Slide 1, and that has a dv of 38.3, which was the worst VAQ in Test 1. Therefore subjects saw the worst case conditions, and then were asked to rank a mid-range slide. They may have over-adjusted their view of "unacceptability" because the second slide shown had such an improvement in visibility over the first slide. Further into the survey, it appears that respondents were more able to calibrate their unacceptability ratings across the entire set of VAQ they were being shown.

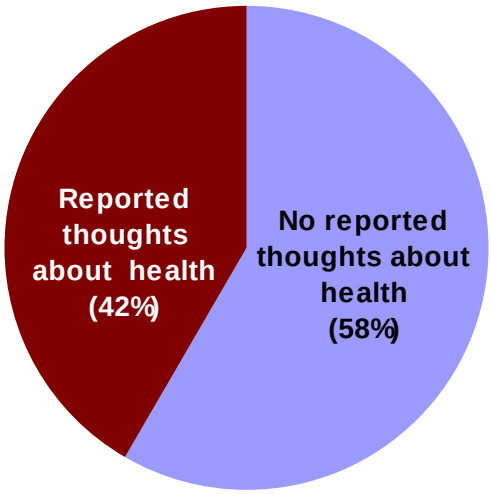
4.3. FOLLOW-UP QUESTIONS

4.3.1. Health Concerns

Once they had finished the survey, respondents were asked, "Was any part of your assessment impacted by thoughts about health?" Figure 10 shows that 42% of the subjects said that they had thought about health, despite having been told by the facilitator prior to rating the slides, "For this discussion, we want you to focus only on visibility: how well you can see. Our goal is to understand your opinions regarding visibility itself. When you give us your opinion, consider only the visibility."

Neither the pilot group nor the Houston group was asked this follow-up question, so the figure below only reflects the responses of the March 2-4 DC respondents.

Figure 10: Evidence of Health Concerns Intertwined in Responses to VAQ Preference Questions



Total Sample of 36 (D.C. Test 1-3 Subjects)

One individual, in response to this question, replied, "But you told me not to!" Many others, however, seemed to have forgotten that they had been instructed not to consider health effects. Many said something along the lines of, "Yes, I was thinking about how hard it would be to breathe or exercise in that air." A number of subjects reported that it was difficult to separate health concerns from the aesthetic effects of low visibility air. The intertwined nature of the two issues was clear from our follow-up questions.

4.3.2. Subject Comments

In the first few pages of the Abt survey, subjects responded to four multiple choice questions while viewing the vista with 15.6 dv, which reflects quite good visibility (see Figure 1). The same questions were posed to the CRA sample. The questions were multiple choice, so subjects chose between discrete options. Figures 11 and 12 show the Houston and DC samples' answers to Questions A and C, respectively.

A number of DC respondents chose to check the "Other" box for Questions A and C. They commented that the survey's definition of "frequently" as "once a day" and "occasionally" as "once a week" were too disparate, and wrote on the survey that they see a view like this, and notice the quality of visibility, a few times a week.

Figure 11: Houston Responses to Questions A and C

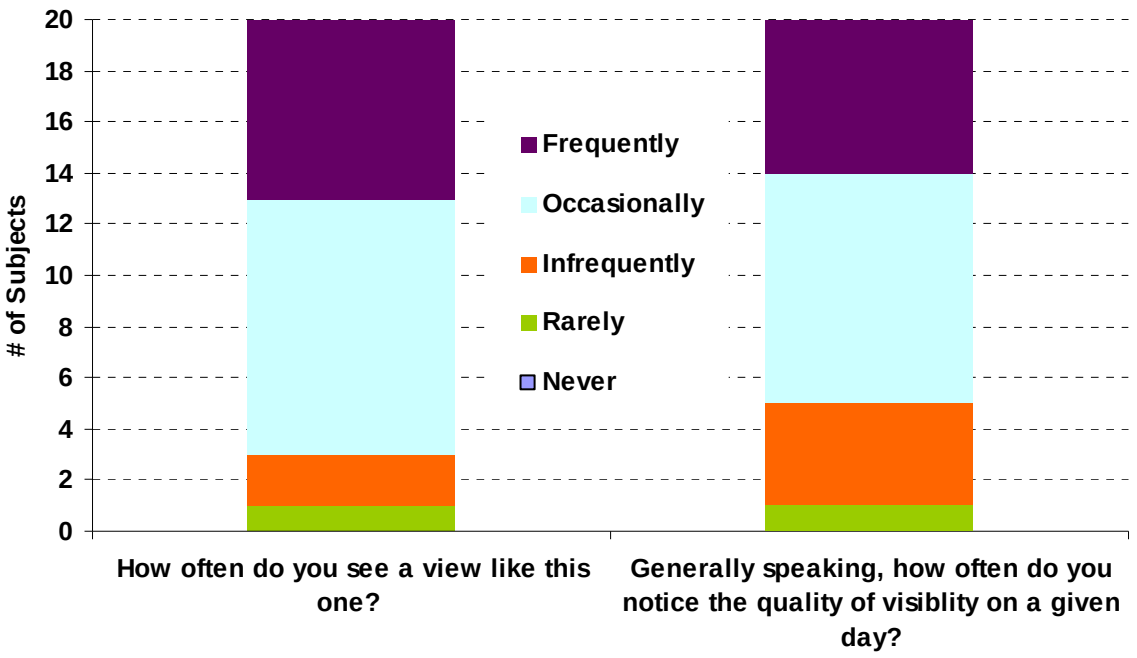
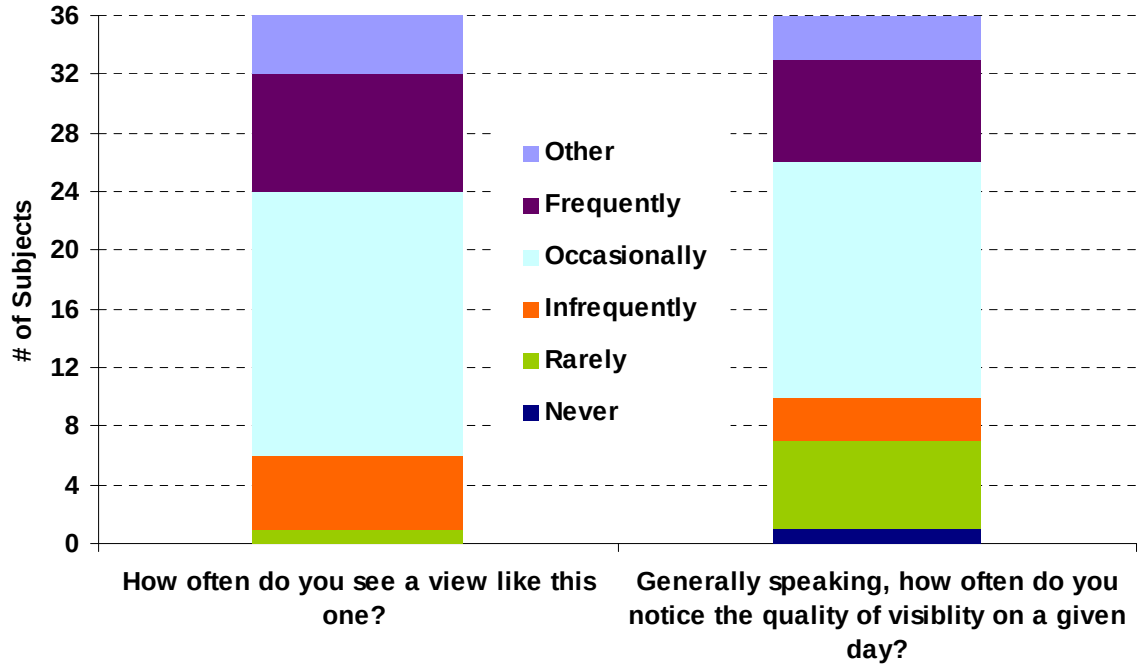


Figure 12: DC Responses to Questions A and C



Question D asked subjects how important visibility is to them. Responses of the Houston and DC samples are shown in Figures 13 and 14, respectively. A number of subjects reflected that "extremely important" was too strong a term. If the survey had used "very important," they would have marked that box, but "extremely important" indicated excessively strong feelings. One respondent said it was an "inflammatory" choice of words. This might explain why relatively few subjects reported that VAQ was "extremely important."

Question B was "Where are you/what are you doing when you see a view like the one on the slide?" The options were "At work," "At home," "Commuting," "Performing day-to-day activities," and "Outdoor Leisure Activities." The facilitator observed that most subjects were confused by this question. They tended to check multiple boxes, or say something along the lines of, "I don't get this." The facilitator felt that it was a poorly designed question and did not find the results worth attempting to summarize here.

In addition to the first follow-up question about health, CRA's subjects were also asked, "In deciding whether a slide was acceptable or unacceptable, what were you thinking about, or what were your criteria?" Most individuals answered that they were considering how hazy it was, how much color they could see, or how far they could see. Some subjects related the vista in the slides to one where they used to live. For example, one participant said, "I'm from Arizona, I was thinking about how clear it is there on a good day and trying to compare." This suggests that some people may attempt to form a view about visibility "acceptability" based on where they grew up or have spent a large amount of time. It is not clear what implication this tendency might have for people's responses to a VAQ preference survey about a location different from that in which they have spent important parts of their lives.

A number of subjects, however, made comments that support our hypothesis that there is no "absolute" standard for acceptable VAQ. Houston Subject #14 said, "I looked at how grey it was, comparing to the best blue I saw in the set, it should be at least above average, not average, to be acceptable." This was supported by comments that subjects made while completing the rating, and that were noted by the facilitator. Houston Subject #16's comments were very much like other subjects' comments. She said, while rating the first slide, "Since I don't have anything to base it on, I'll go with the one in the middle." Then, when viewing slide 2, she said, "This one is better, so I'll go one to the right" (*i.e.*, mark the box that was one degree better than the box she had previously marked).

Figure 13: Houston Responses to Question D

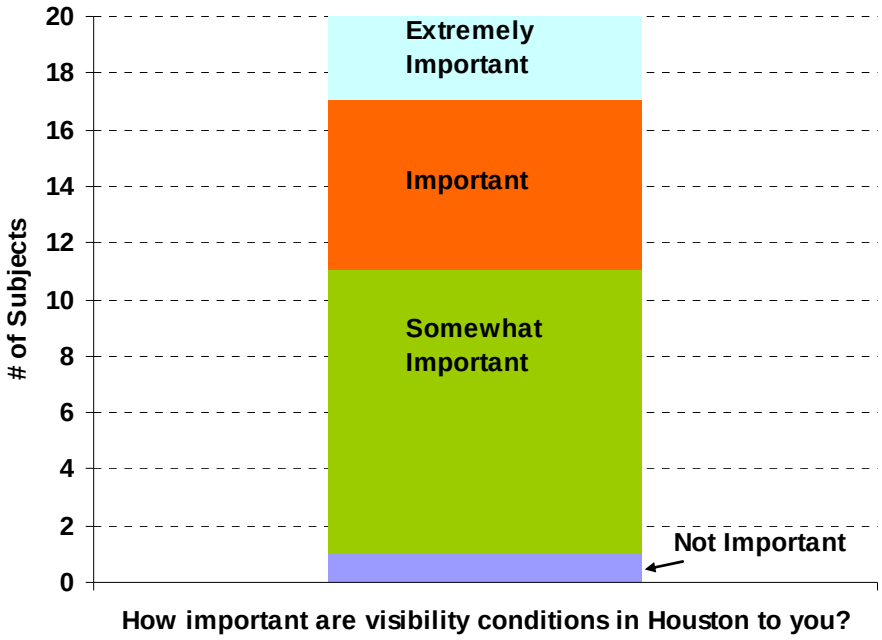
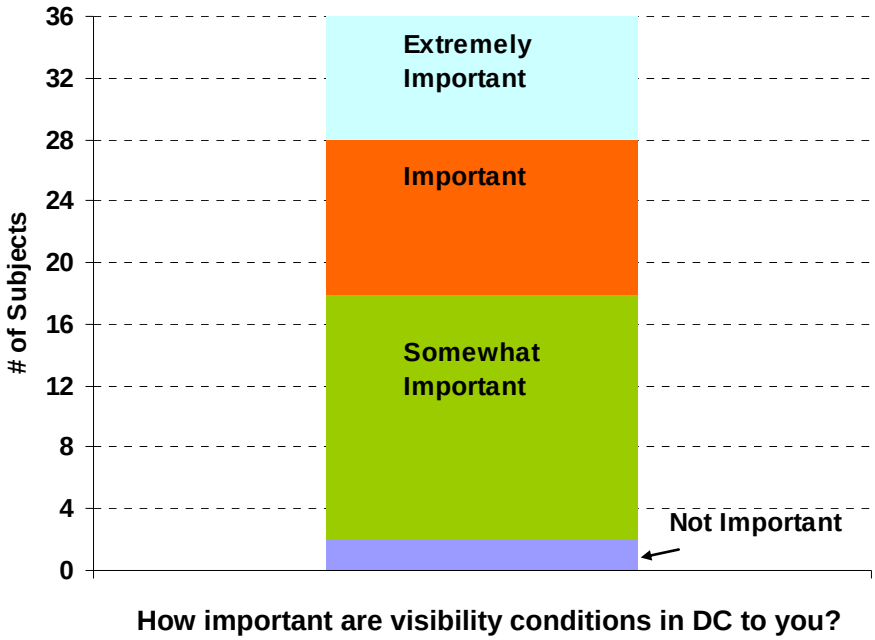


Figure 14: DC Responses to Question D



5. CONCLUSIONS

This survey was conducted on a rapid turnaround schedule by CRA in order to give preliminary exploration to a relatively simple hypothesis about the reliability of VAQ preference studies. As such, it does have a number of limitations, but we believe that these results indicate serious methodological flaws in the VAQ preference study methodology. Our results do, however, need to be confirmed with a more thorough study with larger sample sizes, statistical tests of significances, and a greater degree of representativeness of the population at large.

The key indication of a methodological problem is that the concept of "unacceptable" visibility is malleable, and not an absolute. Statements about VAQ "unacceptability" will vary dramatically depending on the range of VAQ shown. This problem cannot be resolved simply by fine-tuning the range of VAQ shown. As long as the survey results are sensitive to the VAQ shown, then they cannot be interpreted as indicating any reliable indicator of a cutoff below which visibility should not be allowed to go. If VAQ preference studies were to be designed to span the range of currently-experienced VAQ, then this study suggests that they will probably report that the current range of VAQ is unacceptable, even as that range continues to improve to levels deemed acceptable in the previous time period.

In performing further research to see if our findings of sensitivity and lack of robustness are replicable in more representative settings, a number of additional tests should be considered. For example, one could also explore how ordering of the slides (or showing subjects the slides from best to worst or vice versa) might change survey results. Also, scripts that more clearly instruct participants to ignore health concerns, perhaps at multiple points during the survey, ought to be tested.

If our preliminary results are proven replicable, it will be important to find a way to revise the methodology so that it will produce robust results that are not sensitive to the range of VAQ shown in the survey. It is not at all obvious to us that this can be done, however, as it strikes at the heart of the design of this rather simplistic survey technique.

APPENDIX A: KEY SHOWING ORDER IN WHICH EACH OF ABT'S "SLIDE NUMBERS" WAS PRESENTED IN THE QUESTIONNAIRE

Table A-1: Original Key (for Test 1)

Booklet Number	Slide Number
1	Slide 1
2	Slide 16
3	Slide 3
4	Slide 5
5	Slide 22
6	Slide 18
7	Slide 25
8	Slide 11
9	Slide 9
10	Slide 7
11	Slide 13
12	Slide 20
13	Slide 24
14	Slide 12
15	Slide 15
16	Slide 21
17	Slide 14
18	Slide 19
19	Slide 23
20	Slide 17
21	Duplicate 1 / Slide 22
22	Duplicate 2 / Slide 25
23	Duplicate 3 / Slide 23
24	Duplicate 4 / Slide 5
25	Duplicate 5 / Slide 13

APPENDIX B: CRA'S SCRIPT

Each subject received the following script:

1. I first want to find out whether this common degree of clarity that you typically see. Look at slide - this is a pretty typical view of Washington, DC on a day with good weather, and pretty scenic. Now please go to questions 1 and 2 on page 1 of your booklet.

2. Visibility

    a. Air pollutants come from a number of sources, including cars and buses, industry, and so forth.

    b. Air pollution forms smog (also called ground level ozone) and soot (made up of very small airborne particles). Smog and soot in the air cause visibility impairment, or haze.

    c. Natural factors such as humidity and dust can also contribute to haze. Visibility is affected by haze. When it is hazy, you can't see as far, the objects in your view are not as clear and crisp, and colors don't show up as well. Haze is caused when sunlight encounters these tiny particles in the air Some of these tiny particles reduce visibility by absorbing and others reduce visibility by scattering light.

    d. You may also know that the same air pollutants that form haze and reduce visibility are of concern because they are harmful to people and the environment.

    e. For this discussion, we want you to focus only on visibility: how well you can see. Our goal is to understand your opinions regarding visibility itself. When you give us your opinion, consider only the visibility.

    f. Just focus on how things look:

        i. how far you can see

        ii. the clarity

        iii. the crispness

        iv. the colors

3. Rating of Visibility Conditions

a. Now, you will now be shown a series of slides with the same view of Washington we've been looking at. These slides illustrate many different levels of visibility.

b. Just rate the slides according to the visibility conditions in the slides: The ratings run from very poor to very good. Very poor is all the way to the left, and very good is all the way to the right

4. Acceptability/Unacceptability

a. Now you will be shown the same set of slides that you just rated. Again each image will illustrate the effect of a different level of visibility. This time, rate the slides according to whether the visibility is acceptable or unacceptable to you.

5. In deciding whether a slide was acceptable or unacceptable, what were you thinking about, or what were your criteria?

APPENDIX C: DR. SMITH'S EMAIL TO CRA DC STAFF ABOUT CRA'S SURVEY

From: Smith, Anne

Sent: Tuesday, March 03, 2009 1:40 PM

To: *All-DC

Cc: Howell, Sabrina

Subject: Request for 10 minutes of your time

To all of the people in the DC office:

As part of one of my current projects, we are doing a quick replication of a survey that was done in the past by another group. Our goal is just to scope out whether the original results would be likely to be replicable in a more scientifically conducted version of the survey. To this end, we are interviewing only CRA staff who are able to take a few minutes to participate. Sabrina Howell, a CRA analyst, is conducting the survey for me while she is in the DC office this week. I am sending out this message to let you know that she may ask you if you will take the survey, and I hope that you will be willing to help out if she does ask you. The interview, which Sabrina will conduct in the main conference room, takes only about 10 minutes. If you are too busy when Sabrina comes by, perhaps you could arrange to take a 10-minute break later on to do it. If you do participate, your responses will be confidential.

If you have any questions, please let me know. Also, I'll be happy to tell you more about the purpose of the survey once we have completed it. Thanks in advance for you help.

Anne E. Smith
Vice President

CRA International
1201 F Street, N.W., Suite 700
Washington, DC 20004
work: 202.662.3872
mobile: 202.361-4004
email: asmith@crai.com

APPENDIX D: RESPONSE BOOKLET USED IN CRA SURVEY

Subjects were given the exact same booklet as those in the Abt Associates study, a copy of which is in the following pages. There were a few minor alterations, however:

- For the Houston survey, question 2 was crossed out and respondents were told to move directly to question C.
- For the Houston survey, "Washington, DC" was crossed out in question D and replaced with "Houston."
- The Abt survey seems to contain an error in the slide rating section (part E). They have omitted the words "good" and "poor," so the subject has no way of knowing which way the scale works. As seen on page 37-38, this was corrected for the CRA study.

Appendix C: Response Booklet

Washington, DC

November 16, 2000

Name_____

A. How often do you see a view like the one in this slide?
Please CHECK (✓) the appropriate response below.

☐ Never

☐ Rarely (once a year)

☐ Infrequently (once a month)

☐ Occasionally (once a week)

☐ Frequently (once a day)

☐ Other (Please fill in:_____)

2. Where are you / what are you doing when you see a view like the one in the slide? Please CHECK (✓) the appropriate response below. You may check more than one.

- ☐ Commuting to/from work
- ☐ At home
- ☐ At work
- ☐ Performing day-to-day activities
- ☐ Outdoor leisure activities
- ☐ Other (Please fill in:_____)

Abt Associates Inc. Page 37 *January 2001*

C. Generally speaking, how often do you notice the quality of visibility on a given day?
Please CHECK (✓) the appropriate response below.

- ☐ Never
- ☐ Rarely (once a year)
- ☐ Infrequently (once a month)
- ☐ Occasionally (once a week)
- ☐ Frequently (once a day)
- ☐ Other (Please fill in:_____)

D. How important are visibility conditions in the Washington, DC area to you?

Please CHECK (✓) the appropriate response below.

- ☐ Not important at all
- ☐ Somewhat important
- ☐ Important
- ☐ Extremely important

E. Please rate the quality of visibility for each of the images you see using the scale below.
Please CHECK (✓) the response below that most closely reflects your opinion about the visibility conditions in the slide.

Example	Very Poor						Very Good
1	☐	☐	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐	☐	☐
6	☐	☐	☐	☐	☐	☐	☐
7	☐	☐	☐	☐	☐	☐	☐
8	☐	☐	☐	☐	☐	☐	☐
9	☐	☐	☐	☐	☐	☐	☐
10	☐	☐	☐	☐	☐	☐	☐
11	☐	☐	☐	☐	☐	☐	☐
12	☐	☐	☐	☐	☐	☐	☐

Question 5 continued on next page.

Question 5 continued.

	Very Poor						Very Good
13	☐	☐	☐	☐	☐	☐	☐
14	☐	☐	☐	☐	☐	☐	☐
15	☐	☐	☐	☐	☐	☐	☐
16	☐	☐	☐	☐	☐	☐	☐
17	☐	☐	☐	☐	☐	☐	☐
18	☐	☐	☐	☐	☐	☐	☐
19	☐	☐	☐	☐	☐	☐	☐
20	☐	☐	☐	☐	☐	☐	☐
21	☐	☐	☐	☐	☐	☐	☐
22	☐	☐	☐	☐	☐	☐	☐
23	☐	☐	☐	☐	☐	☐	☐
24	☐	☐	☐	☐	☐	☐	☐
25	☐	☐	☐	☐	☐	☐	☐

Abt Associates Inc. Page 42 *January 2001*

F. Please indicate whether you find the quality of visibility Acceptable or Unacceptable for each of the slides presented.
Please CHECK (✓) the box below that most closely reflects your opinion about the slide.

Slide	Unacceptable	Acceptable	Comments
1	☐	☐	
2	☐	☐	
3	☐	☐	
4	☐	☐	
5	☐	☐	
6	☐	☐	
7	☐	☐	
8	☐	☐	
9	☐	☐	
10	☐	☐	
11	☐	☐	
12	☐	☐	

Question 6 continued on next page.

Question 6 continued.

Slide	Unacceptable	Acceptable	Comments
13	☐	☐	
14	☐	☐	
15	☐	☐	
16	☐	☐	
17	☐	☐	
18	☐	☐	
19	☐	☐	
20	☐	☐	
21	☐	☐	
22	☐	☐	
23	☐	☐	
24	☐	☐	
25	☐	☐	

Abt Associates Inc. Page 45 *January 2001*

APPENDIX E: SUBJECT KEY

The tables in this appendix identify the demographic make up of each subgroup in the CRA survey, based on respondent's sex and job position. "EA" stands for Executive Assistant, "ITS" stands for Information Technology Services, and "VP" stands for Vice President.

Table E-1: DC Test 1 Subjects, Interviewed March 2-4, 2009

Subject #	Sex	Position
1	M	ITS
2	M	Senior Consultant
3	F	Analyst
4	M	Senior Consultant
5	M	Analyst
6	F	EA
7	M	Associate Principal
8	F	EA
9	M	Principal
10	M	Principal
11	M	Principal
12	M	EA

Table E-2: DC Test 2 Subjects, Interviewed March 2-4, 2009

Subject #	Sex	Position
13	F	Analyst
14	M	Associate
15	M	Associate
16	M	Associate
17	F	EA
18	M	VP
19	F	EA
20	M	Senior Associate
21	M	VP
22	M	Associate
23	M	Associate Principal
24	M	ITS

Table E-3: DC Test 3 Subjects, Interviewed March 4, 2009

Subject #	Sex	Position
25	M	Associate
26	M	Principal
27	M	VP
28	F	Associate
29	M	VP
30	M	VP
31	F	Principal
32	M	Analyst
33	M	Principal
34	M	Principal
35	M	Analyst
36	M	Associate

Table E-4: Houston Test 1 Subjects, Interviewed February 26, 2009

Subject #	Sex	Position
1	M	VP
2	M	Associate
3	M	Analyst
4	M	VP
5	F	EA
6	M	Analyst
7	M	Office Services
8	M	Principal
9	F	Analyst
10	M	ITS

Table E-5: Houston Test 2 Subjects, Interviewed February 26, 2009

Subject #	Sex	Position
11	F	Analyst
12	M	Associate
13	M	VP
14	M	VP
15	F	EA
16	F	Associate
17	M	Senior Associate
18	F	Office Services
19	F	EA
20	M	VP

APPENDIX F: VAQ SCORES BY CRA RESPONDENT

Note: 1 = Very poor, 7 = Very good

Table F-1: DC Test 1 Ratings of Visual Air Quality

	Subject #											
Slide Number Test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	3	2	2	2	5	3	2	3	3	3	6	3
2	4	5	7	3	7	5	6	2	6	5	7	6
3	2	2	3	1	4	4	3	3	5	2	3	4
4	3	3	2	2	5	3	3	3	6	2	3	4
5	5	7	7	6	7	6	7	5	7	5	7	7
6	4	5	3	5	6	5	5	5	7	5	6	7
7	5	7	7	6	7	6	7	5	6	6	7	7
8	2	3	1	4	4	3	3	3	3	3	4	4
9	2	5	2	4	4	2	3	3	4	3	3	4
10	2	5	1	3	3	2	3	3	3	3	2	3
11	3	6	6	5	5	3	5	4	5	4	2	4
12	4	7	7	6	6	5	6	6	6	6	4	6
13	5	7	7	7	7	5	7	5	6	7	7	7
14	3	3	2	5	4	3	5	3	3	2	4	5
15	4	4	3	5	5	4	5	3	3	3	5	6
16	5	5	6	6	6	5	6	4	5	6	6	7
17	3	4	3	3	5	3	4	3	4	3	5	6
18	4	6	6	4	6	4	5	4	5	5	6	7
19	5	7	7	5	7	6	7	5	6	7	7	7
20	4	4	2	3	6	4	6	3	4	3	4	4
21	6	5	6	5	6	5	6	4	5	7	5	6
22	7	7	7	6	7	6	7	5	6	7	6	7
23	7	7	6	5	6	6	6	4	6	7	5	7
24	2	3	1	2	4	2	1	2	2	2	3	2
25	3	5	3	3	4	3	2	3	3	2	2	3

Table F-2: DC Test 2 Ratings of Visual Air Quality

	Subject #											
Slide Number Test	13	14	15	16	17	18	19	20	21	22	23	24
5	6	5	5	5	7	6	7	7	6	7	3	6
6	6	3	5	4	6	6	6	7	4	7	2	5
7	7	6	6	4	7	6	7	7	5	7	5	7
12	5	3	5	4	6	6	6	7	4	6	2	5
13	6	6	7	5	7	6	7	7	4	7	5	7
16	5	3	6	3	5	6	7	7	3	6	2	4
18	5	2	4	3	5	6	5	6	2	6	1	3
19	6	6	7	6	5	6	6	7	4	7	3	4
20	3	1	4	4	5	5	4	5	3	5	1	2
21	6	4	6	5	6	6	6	6	5	7	2	5
22	6	7	7	6	7	6	7	7	6	7	4	7
23	6	6	7	6	6	5	7	7	5	6	4	6

Table F-3: DC Test 3 Ratings of Visual Air Quality

	Subject #											
Slide Number Test	25	26	27	28	29	30	31	32	33	34	35	36
1	6	5	3	5	6	3	4	6	6	4	6	3
2	5	4	2	4	5	2	3	4	4	3	5	2
3	7	6	4	5	7	4	5	6	7	5	6	5
4	2	2	1	1	3	1	2	1	1	1	2	1
5	7	6	5	7	7	5	6	7	7	7	6	6
6	5	5	4	5	6	3	6	5	5	4	4	4
7	4	4	3	4	4	2	4	4	4	3	4	3
8	5	5	4	4	5	3	5	5	5	4	3	4
9	4	4	2	3	3	2	3	3	4	2	3	2
10	4	3	2	3	3	2	3	4	3	2	3	1
11	3	2	1	2	3	2	2	3	2	1	3	1
12	4	5	3	5	5	4	5	5	5	4	5	3
13	7	6	6	6	6	5	6	7	7	7	6	5
14	2	5	3	3	4	4	3	5	5	4	4	3
15	1	2	1	1	2	1	2	2	1	2	1	1
16	3	4	5	5	4	3	5	5	3	3	5	3
17	3	4	3	4	4	3	5	5	3	4	4	2
18	3	3	2	2	3	2	3	3	2	2	4	2
19	5	6	6	5	5	4	6	6	6	7	6	4

Table F-4: Houston Test 1 Ratings of Visual Air Quality

	Subject #									
Slide Number Test	1	2	3	4	5	6	7	8	9	10
1	7	7	7	7	6	7	7	6	6	5
2	6	7	7	7	6	7	6	5	5	5
3	7	6	7	7	7	7	7	7	6	6
4	4	6	7	6	5	6	5	4	4	5
5	6	6	2	5	5	7	5	4	4	5
6	7	6	6	6	5	7	6	6	5	6
7	2	5	6	6	3	6	4	4	4	4
8	5	6	7	5	3	7	4	5	4	5
9	6	5	5	6	5	7	5	5	4	6
10	3	5	1	3	3	6	3	3	3	4
11	6	5	2	5	4	7	5	5	4	5
12	3	4	1	5	1	5	3	3	1	4
13	5	4	3	5	3	7	3	5	4	5
14	2	5	2	3	1	6	3	3	3	4
15	2	4	1	2	4	5	3	3	2	4
16	3	4	2	3	3	6	3	3	3	2
17	2	4	1	2	2	4	2	3	3	4
18	2	4	1	3	1	5	2	4	2	4
19	2	4	1	3	2	6	3	4	2	4
20	2	4	1	2	2	6	3	4	1	3
21	2	3	1	2	2	6	2	3	1	3
22	2	2	1	4	1	6	4	2	2	4
23	2	3	1	1	1	3	1	2	2	3
24	2	3	1	3	1	5	4	2	1	4
25	2	3	1	3	1	4	2	3	2	4

Table F-5: Houston Test 2 Ratings of Visual Air Quality

	Subject #									
Slide Number Test	11	12	13	14	15	16	17	18	19	20
5	7	6	6	7	7	6	7	6	5	7
6	7	7	7	7	7	5	7	7	5	5
7	7	6	5	6	6	5	7	5	4	5
12	5	5	5	4	6	3	6	5	4	5
13	5	6	5	6	4	4	6	5	4	4
16	5	5	5	5	6	5	5	5	2	4
18	6	6	7	6	7	4	6	6	4	7
19	1	4	4	4	5	3	3	4	3	3
20	4	4	4	4	5	3	3	3	3	3
21	1	3	3	3	5	2	3	2	2	2
22	3	5	6	5	7	3	4	4	3	4
23	2	3	2	4	6	2	2	1	1	2

APPENDIX G: RATING OF VAQ UNACCEPTABILITY BY CRA RESPONDENT

Note: 1 = Acceptable, 2 = Unacceptable

Table G-1: DC Test 1 Ratings of Acceptability/Unacceptability

	Subject #											
Slide Number Test	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	1	2	2	2	2	2	2	2	2	2	2	2
2	1	1	1	1	1	1	1	2	1	1	1	1
3	1	2	2	2	2	2	2	2	2	2	2	2
4	1	2	1	2	2	2	2	2	2	2	2	2
5	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	2	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1	1	1
8	1	2	2	2	2	2	2	2	1	2	2	2
9	1	2	2	2	2	2	2	2	2	2	2	1
10	1	1	2	2	2	2	2	2	2	2	2	2
11	1	1	1	2	1	1	2	2	1	1	2	1
12	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	1	1	1	1	1	1	1	1
14	1	2	2	2	1	2	2	2	1	1	2	2
15	1	1	1	2	1	2	2	2	1	1	2	1
16	1	1	1	1	1	1	1	1	1	1	1	1
17	1	1	1	2	1	2	2	2	1	1	2	1
18	1	1	1	1	1	1	1	2	1	1	2	1
19	1	1	1	1	1	1	1	1	1	1	1	1
20	1	1	1	2	1	1	1	2	1	1	2	1
21	1	1	1	1	1	1	1	1	1	1	1	1
22	1	1	1	1	1	1	1	1	1	1	1	1
23	1	1	1	1	1	1	1	1	1	1	1	1
24	1	2	2	2	2	2	2	2	2	2	2	2
25	1	1	1	2	2	2	2	2	1	2	2	1

Table G-2: DC Test 2 Ratings of Acceptability/Unacceptability

	Subject #											
Slide Number Test	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
5	1	1	1	1	1	1	1	1	1	1	1	1
6	2	2	1	2	1	1	2	2	1	2	2	2
7	1	1	1	1	1	1	1	1	1	1	1	1
12	2	2	1	1	1	1	2	2	1	1	2	2
13	1	1	1	1	1	1	1	1	1	1	1	1
16	2	2	2	2	1	1	2	1	1	1	2	2
18	2	2	2	2	1	2	2	2	1	2	2	2
19	1	1	1	1	1	1	1	1	1	1	2	2
20	2	2	2	2	2	2	2	2	2	2	2	2
21	1	1	1	1	1	1	1	1	1	1	2	1
22	1	1	1	1	1	1	1	1	1	1	1	1
23	2	1	1	1	1	1	1	1	1	1	2	1

Table G-3: DC Test 3 Ratings of Acceptability/Unacceptability

	Subject #											
Slide Number Test	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
1	1	1	1	1	1	1	1	1	1	1	1	2
2	2	2	2	2	2	2	1	2	2	1	2	2
3	1	1	1	1	1	1	1	1	1	1	1	1
4	2	2	2	2	2	2	2	2	2	2	2	2
5	1	1	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1	2	2
7	1	2	2	2	1	2	1	2	2	1	2	2
8	1	1	1	1	1	1	1	1	2	1	1	2
9	2	2	2	2	2	1	1	2	2	1	2	2
10	2	2	2	2	2	2	1	2	2	2	2	2
11	2	2	2	2	2	2	1	2	2	2	2	2
12	1	1	1	1	1	1	1	1	1	1	1	2
13	1	1	1	1	1	1	1	1	1	1	1	1
14	2	1	1	2	1	1	1	1	2	1	1	2
15	2	2	2	2	2	2	1	2	2	2	2	2
16	1	1	1	1	1	1	1	1	1	1	1	2
17	1	1	2	2	1	1	1	1	1	1	1	2
18	2	2	2	2	2	2	1	2	2	2	2	2
19	1	1	1	1	1	1	1	1	1	1	1	1

Table G-4: Houston Test 1 Ratings of Acceptability/Unacceptability

	Subject #									
Slide Number Test	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	1
5	1	1	1	2	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	2	1	1	1
10	1	1	1	1	1	1	2	1	1	1
11	1	1	2	1	1	1	2	1	2	1
12	1	2	2	2	2	1	2	1	2	1
13	1	1	1	1	1	1	2	1	1	1
14	1	2	2	2	1	1	2	1	2	2
15	1	2	2	2	2	1	2	1	2	1
16	1	2	2	2	1	1	2	1	1	1
17	1	2	2	2	2	1	2	1	2	1
18	1	2	2	2	1	1	2	1	2	2
19	1	2	2	2	2	2	2	2	2	2
20	1	2	2	2	2	2	2	2	2	2
21	1	2	2	2	2	2	2	2	2	2
22	1	2	2	2	2	1	2	2	2	2
23	1	2	2	2	2	2	2	2	2	2
24	1	2	2	2	2	2	2	2	2	2
25	1	2	2	2	2	1	2	2	2	2

Table G-5: Houston Test 2 Ratings of Acceptability/Unacceptability

	Subject #									
Slide Number Test	11	12	13	14	15	16	17	18	19	20
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	2	1	2	1	1
13	1	1	1	1	1	1	1	1	1	1
16	1	1	1	1	1	2	1	1	1	1
18	1	1	1	1	1	2	1	2	1	1
19	2	2	1	2	1	2	2	2	2	1
20	2	2	2	2	2	2	2	2	2	1
21	2	2	2	2	2	2	2	2	2	2
22	2	2	2	2	2	2	2	2	2	1
23	2	2	2	2	2	2	2	2	2	2