

# Field Methods

<http://fmx.sagepub.com>

---

## **Triangulation and Mixed Methods Designs: Practicing What We Preach in the Evaluation of an Israel Experience Educational Program**

Charles Kadushin, Shahar Hecht, Theodore Sasson and Leonard Saxe

*Field Methods* 2008; 20; 46

DOI: 10.1177/1525822X07307426

The online version of this article can be found at:  
<http://fmx.sagepub.com/cgi/content/abstract/20/1/46>

---

Published by:

 SAGE Publications

<http://www.sagepublications.com>

**Additional services and information for *Field Methods* can be found at:**

**Email Alerts:** <http://fmx.sagepub.com/cgi/alerts>

**Subscriptions:** <http://fmx.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** (this article cites 7 articles hosted on the SAGE Journals Online and HighWire Press platforms):  
<http://fmx.sagepub.com/cgi/content/refs/20/1/46>

# Triangulation and Mixed Methods Designs: Practicing What We Preach in the Evaluation of an Israel Experience Educational Program

CHARLES KADUSHIN

SHAHAR HECHT

*Brandeis University*

THEODORE SASSON

*Brandeis University and Middlebury College*

LEONARD SAXE

*Brandeis University*

*This article clarifies the role of triangulation in mixed methods research designs. Triangulation is used to describe “severe” statistical tests of correlation between alternative quantitative measurements and is employed to test the accuracy of those measurements. In contrast, the broader concept of mixed methods describes the integration of diverse qualitative and quantitative approaches and is employed to build theory. The concepts are developed and demonstrated in relation to a complex field study that evaluated an Israel experience educational program.*

**Keywords:** *mixed methods; triangulation; field methods; evaluation*

Interest in mixed methods in applied research contexts has increased substantially. As a measure of interest and the emergence of a field, there is a new journal (the *Journal of Mixed Methods*) and a recent handbook of mixed methods (Tashakkori and Teddlie 2003). Nonetheless, the logic of mixed methods as they are used by researchers remains unclear, and the lack of clarity limits the application of these strategies. In a recent commentary about the new handbook, Miller and Gatta (2006:596) observe that “MM [mixed methods] as a research enterprise is at this time still unsure as to what it is supposed to accomplish . . . it is not exactly clear how MM are to be used to enhance validity, reliability or deeper interpretations of what is being studied.”

---

*The authors gratefully acknowledge the helpful comments of Barry Chazan, the anonymous reviewers of this article, and the editor, H. R. Bernard.*

*Field Methods*, Vol. 20, No. 1, February 2008 46-65

DOI: 10.1177/1525822X07307426

© 2008 Sage Publications

The difficulty with mixed methods approaches, according to these authors, stems in part from confusion over the specific contribution of triangulation in the research process. Both mixed methods and triangulation are forms of “diverse testing,” but there has never been clarity over what triangulation is supposed to contribute as a method of diverse testing, and as a consequence, triangulation has “not . . . lived up to its expectations” (Miller and Gatta (2006:597). At the root of the problem is the absence of guidelines for the correct application and interpretation of triangulation in mixed methods research. Our aim in this article is to clarify the proper role of triangulation in mixed methods research designs.

Drawing on an evaluation study of Israel experience tours (see Saxe, Sasson, and Hecht 2006), this article will model and demonstrate the discrete contribution of triangulation to mixed methods applied social research. In the literature, as Miller and Gatta (2006:598) observe, triangulation has been used to denote any comparison between methods—usually between qualitative and quantitative methods or the reverse. The present suggestion is that the term *triangulation* be reserved for a specific category of diverse tests—“severe” quantitative tests. Triangulation is thus defined as measuring the same concept using two or more methods. Indicators of the concept are first collected in each method, and the underlying or latent concept is measured within each method (Henry and Lazarsfeld 1968). The measurements taken within each method are then correlated. The goal is that of a more accurate measure; that is, triangulations will achieve a better estimate of the error inherent in any measurement both within and between the methods.

Studies that compare quantitative and qualitative observations at the conceptual or propositional level, as opposed to those that conduct statistical comparisons, should more properly be described as mixed methods studies. A mixed methods design excels at bringing insights derived from diverse methods to the analysis of a given phenomenon. In this endeavor, the indicators themselves, such as field notes and responses to survey questions, may be examined and compared across the different methods and result in the development of new concepts. Mixed methods are therefore central to the development and testing of theory.<sup>1</sup> This is certainly not a new concept, and sophisticated social researchers have routinely used both field and survey methods for years (Sieber 1973).

In our quantitative dominant sequential mixed methods design (Creswell et al. 2003; Hanson et al. 2005; Ivankova, Creswell, and Stick 2006), triangulation constitutes the first phase of the analysis. It is utilized to estimate the error inherent in our quantitative measurements. Although triangulation contributes directly to the analysis of the phenomenon being measured, it also serves as a building block for the incorporation of qualitative data. In

the second phase, the qualitative data are examined to further verify and then deepen the analysis of the quantitative findings. It is hoped that a better understanding of the concepts of triangulation and mixed methods and how they can be employed in research design will facilitate diffusion of mixed methods research.

## ISRAEL EXPERIENCE TOURS

The substantive context for the present analyses is a program that provides young Jewish adults with opportunities to learn about their heritage by participating in educational trips to Israel (see Saxe, Sasson, and Hecht 2006). Although Jewish pilgrimages to the Holy Land have taken place at least since the first century of the Common Era, current Jewish youth trips to Israel have a special flavor. As Kelner (2002) points out:

The contemporary youth pilgrimage-tour in Israel that has developed since the 1950s is less the continuation of an ancient Jewish tradition of pilgrimage than the institution of a new form of Jewish experience and expression, made possible by modern organization, modern technology and modern conceptions of travel and leisure. It exists at the intersection of Israel's tourism industry and Diaspora Jewry's educational system. As the field has institutionalized, it has developed an increasingly explicit conception of itself as an instrument of informal Jewish education, and a deepening conviction that participation in Israel experience programs yields demonstrable effects on Jewish identity. (p. 77)

A variety of methods have been used to study aspects of youth pilgrimage tours to Israel (Goldberg, Heilman, and Kirshenblatt-Gimblett 2002). Goldwater (2002) used participant observations accompanied by video and audio recording when examining the work done by the tour educators. A series of studies, focusing on the long-term impact of the program, used online surveys of participants and nonparticipants (Saxe et al. 2004). Although a small number of studies have used both qualitative and quantitative methods (Saxe et al. 2002), none has examined the methodological issues that arise in mixed methods design.

## STUDY DESIGN

The work reported on here is part of a program of research designed to evaluate the impact of educational trips organized by Taglit-birthright israel (birthright israel). birthright israel does not run the trips itself. Rather, it contracts with a number of tour organizers to bring college-age adults and

recent college graduates to Israel for ten days. The long-term evaluation of birthright israel seeks to identify the key ingredients in successful trips and track trip effects over a long period. The immediate goals of the present study included learning how the tours vary in terms of thematic content, group dynamics, guiding, participant satisfaction, learning, and overall success. Narrative descriptions of the various programs as well as the perspectives of program participants were needed. It was also necessary to identify the combination of methods to utilize in future evaluations of the program. A mixed methods protocol was needed to accomplish these diverse goals.

The birthright israel tours studied as part of this investigation took place in the late spring and early summer of 2005. Although each tour organizer was responsible for as many as two thousand participants, the actual touring took place on individual buses with no more than forty participants who remained with one another throughout the ten-day trip. The bus is, therefore, the unit of observation. A set of twenty-two buses was selected from all one hundred thirty-nine North American buses operating during the 2005 summer session of birthright israel.<sup>2</sup> The selected set of buses was designed to represent all large trip organizers and a select number of smaller organizers. In addition, the buses were selected so as to avoid overlap with buses sampled for the program evaluation by the Israeli evaluator.<sup>3</sup>

### Methods of Data Collection

Our mixed methods research design combined four types of data: observer quantitative ratings on guide performance and group dynamics, observer qualitative narratives on program content, participant observer diaries, and participant questionnaires. Each of these methods is detailed below.

*Observer quantitative ratings and qualitative narratives.* The bus observation protocol called for a rating by the observers on a 5-point scale of a number of dimensions related to the performance of the guide and group atmosphere. These dimensions were also covered in the participants' questionnaire described below. Ratings were to be made during or immediately after the observations. Observers were also asked for an extensive narrative in the more typical fieldwork observation mode. In the observation protocols, observers were asked to describe the different formal components of the program (e.g., site visits, lectures, discussions, organized encounters with Israelis, and other activities) as well as the informal moments, religious rituals, and the general atmosphere on the bus. The protocol required that individual bus observations take place over a full day (breakfast through the evening program conclusion). Each sampled bus was observed for two full

days by the same observer. The first observation took place on days 3–5 of the trip, and the second observation took place on days 7–10 of the trip. In addition, three tour sites that represent major themes relating to Judaism and Zionism (central goals of the program) were selected for additional observation: Masada, the Western Wall and Jerusalem's Old City, and Har Herzl (the military cemetery in Jerusalem where major political leaders, including Theodor Herzl, are buried). Observers stationed at the sites joined the group for the entire activity at the site. In most cases, this amounted to half a day of observation. Site observations were conducted on twelve buses. Observers conducting site observations were usually not the ones who joined the bus for the full two days of bus observations.

The observers included nine individuals, six employed by the Israeli research firm and three hired directly by the present authors. Most of the observers were graduate students in the social sciences, including several with substantial experience in ethnographic field methods. All were fluent in both Hebrew and English. A training session was conducted in Israel for all observers. Observers were instructed on how to introduce themselves to the group, how to use protocols, and how to be an effective note taker. Observers remained in close contact with the research group throughout the summer session. Feedback was provided on all reports submitted, and revisions were made by observers when necessary.

*Participant observations.* Participant observers were present on six of the sample buses. Participant observers were birthright-eligible candidates who went on the trips and kept journals and notes describing their experiences. They were instructed to focus on the educational content of the program. In addition, they were asked to report on activities and conversations taking place during informal moments as well as on the general orientation and atmosphere in the group.

The participant observer team included four undergraduate students, one graduate student, and one college graduate professional. All had taken college-level research methods classes and were familiar with qualitative field methods. Participant observers were instructed on how to balance their role as both participant and observer and be an effective note taker.

*Participant questionnaires.* A participant questionnaire developed for this study was designed to gather participant reactions to the tour. Questions covered the respondents' attitudes toward their traveling group, their guides and trip leaders, the various tour sites, and the experience as a whole. Questionnaires were distributed to participants on the ninth or tenth day of the trip.

## ANALYSIS

The first phase of data analysis triangulated two methods and sources of data: Participants' questionnaire responses were compared with observers' quantitative ratings. Because the comparison was of alternative sets of quantitative measures of the same phenomena (e.g., guide performance, group cohesiveness), the analysis consisted of statistical correlation. In the second phase, narratives developed from field notes and participant diaries were combined with insights derived from the quantitative analyses to generate hypotheses about the ingredients of a successful trip. These two phases of data analysis are discussed below.

### Triangulation: Comparing the Quantitative Measures

The participant questionnaires included seventy-three questions relating to different aspects of the trip, but they can be grouped into a smaller number of scales reflecting different dimensions of the trip experience. Responses to all original questions on the survey were measured on a 5-point scale: a score of 1 representing a low or negative rating and a score of 5 representing a high or positive rating.<sup>4</sup> As a result of a series of principal component, varimax rotation factor analyses, twelve scales were constructed. Typical average loading on items selected for the scales was .7.<sup>5</sup> The twelve scales are described in Table 1. The names assigned to these scales represent the authors' interpretation of the basic sense of the items included; that is, the assigned names are the concepts; the items themselves are the indicators. The scores, based on 837 respondents with a mode of 37 per bus were aggregated by bus. That yielded 22 as the effective number for comparison with the observer ratings.<sup>6</sup>

Objective data from observer protocols focused on three components: the group, the guide(s), and the trip leaders. Data presented here are derived primarily from the overall assessment of the guides and from the assessment of their performance at the sites visited. All references to the guide(s) are to the regular guide(s) that conducted the tour.<sup>7</sup>

Observer ratings were made on a 5-point scale: A score of 1 represented a low or negative rating, and a score of 5 represented a high or positive rating. A factor analysis of the mean scores of the observer ratings by bus was performed, yielding two basic ratings: the general effectiveness of the guide and the quality of the guide's presentations at the sites. The two scales are described in Table 2. Again, the names or the concepts are ours—the ratings are the indicators.<sup>8</sup>

Scores obtained by the two methods were compared, and Table 3 shows Pearson correlations between the two observer rating factor scores and the

TABLE I  
Survey Scales, Items, Means, Standard Deviation, and N

Scale	Survey Items Included	Mean	Standard Deviation	N
I. Ratings of the quality and nature of the bus group				
1. Good group vibes	Extent felt like part of community formed on bus; extent formed friendships with others on bus; people in group felt like group of friends; people in group felt like a community	4.1284	.32493	22
2. Negative group vibes	People in group were disruptive or distracting; extent felt uncomfortable with some people on the bus; extent felt group atmosphere detracted from experience; people in group were disrespectful toward each other; people in group formed separate cliques	2.5494	.35498	22
3. Similar observance	Degree to which other people on bus compared to me in terms of Jewish observance; degree to which trip compared to me in terms of Shabbat observance	3.4790	.24368	22
4. Partying atmosphere	People in group were on trip just for good time; people in group were not on trip to explore Jewish identity	2.8229	.36583	22
II. Extent and nature of learning on the trip and the nature of the experiences				
5. Learned about Israel and Judaism	Learned about Jewish history; learned about Jewish customs and practices; learned about Israeli culture; learned about Israeli social problems; learned about Israel's landscape	3.9434	.23011	22
6. Jewish religious experience	Thought about increasing religious observance, experienced moments of religious inspiration, thought about self differently as Jew	3.2293	.34566	22
7. Learned about Arab-Israeli conflict	Perspective on Arab-Israeli conflict was reinforced; given balanced view of Israeli-Palestinian conflict; learned about Israeli-Palestinian conflict	3.2979	.24317	22
8. Interest in learning more about Israel	Wish had more time in Israel; want to learn more about Israel	4.2379	.19489	22



III. The guides and trip leaders			
9. Guide open and available	Israeli tour guides friendly to everyone on trip; Israeli tour guides open to concerns or questions; Israeli tour guides contributed to feeling of community on bus; Israeli tour guides available to talk about trip experiences; Israeli tour guides sensitive to needs of group; Israeli tour guides encouraged everyone to participate fully	4.3825 4.4647	.30362 .38993
10. Guide expert and understandable	Israeli tour guides were easy to understand; Israeli tour guides were knowledgeable		22 22
11. North American trip leaders' ties to group	Trip leaders friendly to everyone on trip; trip leaders open to concerns or questions; trip leaders contributed to feeling of community on bus; trip leaders available to talk about trip experiences; trip leaders sensitive to needs of group; trip leaders encouraged everyone to participate fully; trip leaders knowledgeable	4.2389	.37238
12. Political and religious preaching	Israeli tour guides promoted one religious perspective; trip leaders promoted one religious perspective; Israeli tour guides promoted one political perspective on Israel; trip leaders promoted one political perspective on Israel	2.7784	.36129

---

**TABLE 2**  
**Observer Scales and Ratings, Means, Standard Deviation, and N**

<i>Scale</i>	<i>Observer Rating Scales Included</i>	<i>Mean</i>	<i>Standard Deviation</i>	<i>N</i>
The general effectiveness of the guide	The tour guide's command of the language; the guide's control of the group in terms of holding group's attention; the guide's sensitivity to the group's physical state; the guide's sensitivity to the group's emotional state; the group's general level of interest	3.9089	.60827	22
Site presentation quality	The tour guide's knowledge of the sites; the guide's quality of the presentation ( <i>boring to riveting</i> )	3.9476	.70002	21

twelve factor scores of respondents, both aggregated by bus. The number of cases is twenty-two for the "general effectiveness of guide" score and twenty-one for the "site presentation quality" score.<sup>9</sup> Seven of the twelve correlations between the observers' general effectiveness score and the respondents' scores are significant at least at the .05 level (shown in bold type). Three of the site presentation quality scores are significant at the .05 level or better. The significant findings are not a result of the number of tests conducted; for example, given that seven correlations between participant scales and the observers' scales are significant at the .03 level, the chance of finding seven of these in twelve observations is not statistically significant (Bonferroni adjustment is .22).

The average absolute value of the correlations for general effectiveness of guide, including all twelve, is .438 accounting for 19% of the variance. The error is large relative to what might be expected in a laboratory, but typical of correlations in field and survey data.<sup>10</sup> Nonetheless, one gets a sense of what makes for an effective guide: to be open and available and to be expert and understandable. Observers also rated bus groups on a variety of measures of cohesion and orientation toward the program. The correlation between the observers' scores for group cohesion and the participant questionnaire factor for "good vibes" is .450 ( $p < .05$ ). This finding reinforces the notion of what a good guide might accomplish. The advantage of numerically assessing triangulation rather than simply declaring that results look similar is apparent. The findings reinforce the idea that key concepts measured in the field observations correlate with key concepts measured in the questionnaire. Nevertheless, if one wants to compare all of the concepts, a Bonferroni adjustment shows that a statistical comparison of all the concepts in the

TABLE 3  
Key Participant and Observer Scores Compared (Pearson's *r*)

<i>Participants' Scales</i>	<i>Observers' Scale Scores</i>	
	<i>General Effectiveness of Guide</i>	<i>Site Presentation Quality</i>
Good group vibes	.492*	.565*
Sig.	.020	.008
<i>N</i>	22	21
Negative group vibes	-.381	-.393
Sig.	.080	.078
<i>N</i>	22	21
Similar observance	.323	.334
Sig.	.143	.139
<i>N</i>	22	21
Partying atmosphere	-.498*	-.390
Sig.	.018	.080
<i>N</i>	22	21
Learning about Israel and Judaism	.451*	.292
Sig.	.035	.199
<i>N</i>	22	21
Jewish religious experience	.468*	.295
Sig.	.028	.194
<i>N</i>	22	21
Learning about Arab-Israeli Conflict	.216	.098
Sig.	.335	.671
<i>N</i>	22	21
Interest in learning more about Israel	.324	.121
Sig.	.141	.602
<i>N</i>	22	21
North American trips leaders' ties to group	.514*	.160
Sig.	.014	.489
<i>N</i>	22	21
Guide open and available	.617*	.463*
Sig.	.002	.035
<i>N</i>	22	21
Guide expert and understandable	.587*	.628*
Sig.	.004	.002
<i>N</i>	22	21
Political and religious preaching	-.388	-.168
Sig.	.075	.468
<i>N</i>	22	21

\*Correlation is significant at least at the .05 level (two-tailed).

fieldwork with all of the concepts in the survey is not significant. This suggests that taken as a whole, the surveys may not substitute for the field observations since the gestalt of the two is different. We review the wisdom of performing these adjustments in our discussion section.

### Mixed Methods: Bringing in the Qualitative Data

In the second phase of the analysis, the distribution of questionnaire results by bus are compared with the observers' and the participant observers' narratives. Here, we are comparing the indicators as well as the concepts. The comparison is qualitative in that our goal is not to find a statistical correlation; rather, the goal is to assess the basic picture or storyline regarding each bus across the three methods. In addition to further corroborating the questionnaire results, this enables us to develop hypotheses about possible reasons for the key numerical associations in the quantitative survey data. Thus, it allows us to take full advantage of the mixed methods approach.

First, let us review some of the quantitative results. In Table 4, analyses of the scales developed from the participants' questionnaires are summarized. For these analyses, participants are treated as survey respondents rather than as informants about what takes place on their particular bus. As shown in Table 4, the questionnaire results for the performance of the guides, the good vibes of the group, and learning about Israel and Judaism are related. The correlations are statistically significant and typical of the range for survey data. Participants who felt that they had an expert and understandable guide tended to think the guide was open and available. This also helped them learn more about Israel and to experience good group vibes (though there were obviously other sources for this learning, and judging from the size of the correlation, other factors contributed to good vibes). Previously, in Table 3, taking the scores one bus at a time, we saw that bus groups whose guides received top marks (the majority of buses) tended to be good spirited and cohesive, and their participants reported learning a great deal. In contrast, buses whose guides received comparatively lower evaluations—between one-fourth and one-third of the sample—were more fractious, and their participants more divided over the extent of their learning.

We now examine the qualitative data. The narratives prepared by the observers and participant-observers were analyzed to see what light they might shed on these associations. Bus-level data on the performance of the guides are shown in Figures 1 and 2, which display respondents' evaluations of guides across the sample buses. For each bus, the strong solid line shows the median on our scale of 1–5. The solid color, the box, shows where 50% of the values are located (the interquartile range). The horizontal lines,

TABLE 4  
Participant Scales for Guides, Group Atmosphere, and Learning (Pearson's *r*)

	<i>Expert and Understandable Guide</i>	<i>Open and Available Guide</i>	<i>Learn More about Israel</i>	<i>Good Group Vibes</i>
Expert and understandable guide	1	.405**	.131**	.155**
Sig. (two-tailed)		.000	.000	.000
<i>N</i>	828	828	825	828
Open and available guide		1	.228**	.331**
Sig. (two-tailed)			.000	.000
<i>N</i>		833	830	833
Learn more about Israel			1	.374**
Sig. (two-tailed)				.000
<i>N</i>			832	832
Good group vibes				1
Sig. (two-tailed)				
<i>N</i>				837

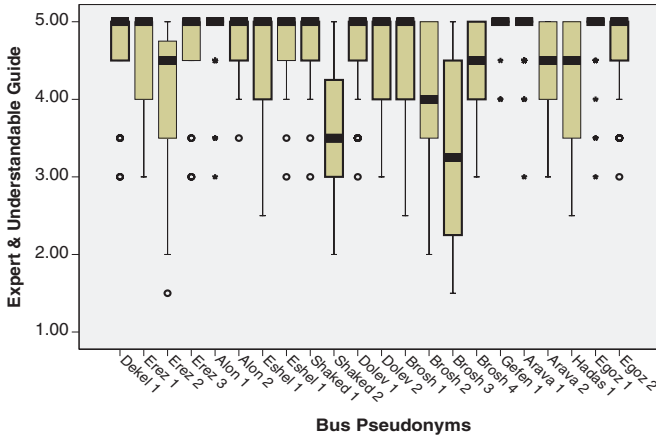
\*\*Correlation is significant at the .01 level (two-tailed).

attached to “whiskers” that extend from the box, show the highest and lowest values, excluding outliers. The hollow dots indicate the locations of outliers (defined as cases with values between 1.5 and 3 box lengths from the upper or lower edge of the box). Asterisks (\*) represent extreme cases, with values more than 3 box lengths from the upper or lower edge of the box.

Most buses scored very well; on the whole, birthright israel runs a very successful program. Six buses, however, scored comparatively lower than the rest. These six included Hadas 1, Shaked 2, Erez 2, and three Brosh buses (2, 3, and 4).<sup>11</sup> The two Alon guides received mixed scores, high for expertise but relatively low for openness to the group. The strongest guides served Gefen 1, Eshel 2, Egoz 1 and 2, and Dekel 1.

Drawing on field reports and participant diaries, a summary narrative was developed for each bus in the sample. The summary bus narratives were then reviewed to try to explain differences in the quantitative ratings across buses in the sample. The summary narrative accounts—especially those sections focusing on the guides—enabled understanding of some of the quantitative differences in ratings across buses. In general, the narratives suggest that the comparatively weaker guides exhibited one or more of the following tendencies: They spoke English less fluently, they had difficulty holding the attention of the group, and they demonstrated less interest in or capacity for connecting personally with tour participants. The following extracts from the observer field reports are illustrative:

FIGURE 1  
Expert and Understandable Guides



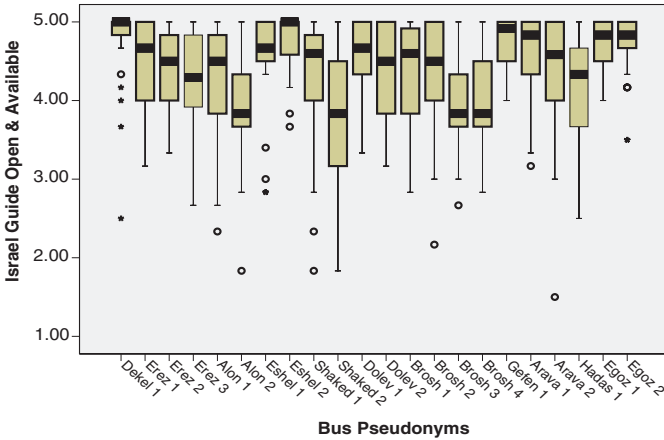
Yonatan’s<sup>12</sup> lack of English language skills is a real problem. Even before I had heard him, a participant whom I spoke to told me that they often can’t figure out what he’s saying. When I heard him speak, the problem was immediately evident. It’s not impossible to understand him; in one-on-one conversation, he is able to communicate, but one has to make an effort. . . . I think around half of what he said was lost (especially educational content regarding sites). (Shaked 2, first observation)

Yoav gathered everyone near Jaffa gate and started to talk about the history of the place. He was shouting throughout all of his guidance and speaking with quite a limited vocabulary and a heavy Israeli accent, which often did not allow him to explain himself properly. When we walked along the walls through the narrow path that surround the walls, he did not pay attention to the last participants and a huge gap became evident between him at the lead of the group and those in the back. Also, one participant became sunburned and asked Yoav for sunscreen, but he said he would only give it to him later when we stop. . . . I might add here also that when talking to some members of the group about the trip, they said that “Yoav does not shut up.” “Even during the bus ride, he always shouts,” and so they cannot really sleep. (Brosh 4, first observation)

In contrast, the stronger guides had most of the skills that the weaker guides were lacking:

Yael’s big advantages are her social skills and her dedication. She really cares about the participants, and as she said in the closing discussion . . .

FIGURE 2  
Open and Available Guides



she gets excited together with the participants. She mentioned her private discussions with participants, and I understood that she stayed up at nights talking to them. This facilitated group cohesion and spirit. (Dekel 1, second observation)

His English was excellent, with minor exceptions which did not seem to interfere with the group understanding his message. . . . The guide’s consistent pleasant and even-minded, even-toned style assumedly plays its role in maintaining the group temperament as it was. . . . The side conversations were rare, and the guide was clearly sensitive to the appropriate measure of information to provide so as not to lose their attention. (Egoz 2, first observation)

The narrative data enable further examination of the statistical relationship among the key scales in the participant questionnaire: good guiding, group vibes, and group learning. In general, buses with top scoring guides performed well on the good vibes and learning scales; those with relatively weaker guides scored lower. The observer ratings and narrative accounts shed considerable light on the dynamics in these buses and provide a fairly clear picture of the successful (and not so successful) trip. On the majority of the buses, guides held the attention of the participants, and the latter reported learning a “great deal” about Israel and Judaism. On the handful of comparatively weaker buses, the guides failed to hold the attention of many of the participants; some listened, while others wandered off. As a

result, in these groups, the participants lacked a common focal experience and felt less a part of a shared group experience. In these buses, moreover, antagonism sometimes developed between participants who attempted to pay attention in spite of the guide's shortcomings and those who preferred to make cell phone calls or chat in the background.

These dynamics are documented by observer reports. For example, the observer report for Alon 1 describes the group's visit on the last day of the tour to Mount Herzl. According to the observer, the guide spoke English with difficulty and was incapable of holding the attention of her group. Fewer than ten participants listened to her remarks at the grave of Hanah Senesh, a Jewish war hero killed behind Nazi lines during World War II: "The rest of the group were wandering all over; some were talking on their cell phones." It is easy to see how this disarray, rooted in ineffective guiding, can generate tensions within the group and an overall sense that more could be learned. In this climate, it is not surprising that participants were less confident in their learning.

These observations served as the basis for our working hypothesis that good guiding fosters trip success by creating a common focus and shared group experience. We refer to this as the "focal experience" hypothesis.<sup>13</sup>

The observers' ratings for group cohesion during the first and second observations provide a test of the focal experience hypothesis. Three buses, Brosh 2, Brosh 3, and Alon 1, were ranked by observers more highly for cohesiveness on the first rather than on the second day of observation.<sup>14</sup> Shaked 2 and Hadas 1 were described as cliquish during both observations. All of these groups were heterogeneous in terms of age and other background characteristics of their participants. What they shared were guides whose performance was below average both at the tour sites and in terms of overall management of the tour group. The opposite tendency can be seen in Eshel 2 and Dolev 2, two groups that became more cohesive between the first and second observations in the context of more effective guides. Examining the observer narratives yields a plausible social psychological explanation: Good leadership ensures communal experiences that bind the group; weak leadership produces disarray and disorientation.

## DISCUSSION

The current study suggests the value of triangulation and its place in mixed methods research and analysis. The triangulation component of the study examined the consistency of the quantitative measures between the



survey and the field observations from a severe point of view (Miller and Gatta 2006). Although key individual correlations between bus scores derived from participants and those derived from field observers are significant as well as substantively illuminating, if we take into account that we are examining an entire set of correlations, using Bonferroni adjustments for the multiple tests, the results as a whole are not significant. Perneger (1998), however, argues that adjusting all the  $p$  values “for the number of tests that have been performed on study data—the Bonferroni method—creates more problems than it solves”. This is so insofar as researchers are often interested in particular concepts. Type II errors (rejecting a “true” hypothesis) are increased with Bonferroni adjustments and “[t]hus, contrary to what some researchers believe, Bonferroni adjustments do not guarantee a ‘prudent’ interpretation of results” (Perneger 1998:1237).

The key correlations between participants’ ratings of their guides as “open and available” (.62) and as “expert and understandable” (.59) and of observers’ ratings of the general effectiveness of the guides build confidence in the validity of our quantitative measures of guide performance. With regard to measures of additional features of the various tours, the correlations are weaker, and the implication seems to be that both questionnaires and fieldwork are needed to provide a complete picture. Future evaluation studies will need to balance the desirability of obtaining as complete a picture as possible against the added cost of conducting field observations alongside a survey.

More broadly, triangulation was used as part of mixed methods research design that aimed not only to assess outcomes but also to build theory. Qualitative comparison of the narrative and quantitative data suggested explanations for the variations in the ratings of different buses that we might have missed otherwise—especially in relation to the significance of focal experiences. Thus, whereas triangulation permitted a sober assessment of the reliability of our key quantitative measurements, the study’s qualitative component facilitated development of our explanatory theory, that is, the meaning of the quantitative findings. This is a good use of sequential mixed methods.

Our conclusion is that strict triangulation should be used when possible as a statistical assessment of the extent to which multiple measurements of key concepts converge. Although often advocated, it is rarely done in practice. Yet, strict triangulation does not advance research. The term *mixed methods*, more broadly, is best reserved for conceptual comparisons of narratives and detailed examination of indicators across different methods, thereby developing more plausible interpretations of the findings in each of the methods.

The present study is limited in several ways. Although it was designed as a quantitative triangulation study to simultaneously evaluate the utility of both the survey and the field observations, the very fact that it was a pilot study restricted the number of units (in this case, buses) that could be studied as well as the time available to develop an evaluation and recommendation. In addition to the precoded observations of the fieldworkers, it would have been preferable to code the observers' field notes and quantitatively compare all three sources of data. The need to quickly arrive at a recommendation for future data collection precluded this additional step. Nevertheless, the qualitative judgment of the meaning of the summarized narratives in conjunction with the quantitative triangulation lent an air of veracity to the recommendations.

These limitations suggest why true quantitative triangulation is slighted in the discussions of mixed methods. First, it is expensive—at least it is perceived to be so: It requires a number of observers—a rarity in fieldwork—a survey or other quantitative data gathering, and systematic precoding of observers' findings as well as systematic after-the-fact coding of their narratives. Second, both the surveys and the precoding of observations imply that one knows what one is looking for. But the underlying rationale for fieldwork is that one is not sure of exactly what will be found in the field, and the prospect of serendipitous findings and interpretations drives the motivation for this arduous and expensive method.

Because something is difficult does not mean that it should not be attempted. In the present case, because we were developing an approach that could be applied more widely (to future cohorts who participate in similar educational efforts), it was cost-effective to test different methods and to examine systematically the relationship across measures. Too often, research strategies are based on whatever methods and strategies the investigator, by dint of history or training, decides to privilege. If we want the findings of applied research to be paid significant attention, it is our obligation not only to approach the problem using different methods, but to triangulate our measures. We need, perhaps, to be as rigorous in comparing our data as we were in collecting the data originally.

## NOTES

1. On mixed methods research and theory building, see Creswell et al. (2003), Hanson et al. (2005), and Johnson and Onwegbuzie (2004). For a dissent arguing that qualitative and quantitative methods represent different worldview and different paradigms, see Sale, Lohfeld, and Brazil (2002). We reject this view.

2. One of the buses selected had a small group of South American participants who were not given the participant questionnaires.

3. The present study was originally conceived as a joint exercise with an Israeli research organization that had devised observation protocols to evaluate the quality of services delivered by the various trip organizers and had been doing so for some years. Large numbers of buses in each trip season were evaluated. The original observation protocols and methods were starting points for the design of the current observational work, but considerable changes were incorporated both to rating scales and narrative reporting as well as to the schedule of observations.

4. The survey included two question formats: The first had a question stem beginning with "To what extent . . ." with a response scale ranging from 1 (*not at all*) to 5 (*a great extent*). The second had a response scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

5. Although many analysts report Cronbach's alpha, these do not guarantee unidimensionality because they are based on manifest rather than latent scores. High loadings on selected items are a much better indicator of unidimensionality because they are based on the latent score (Bond and Fox 2001).

6. Response rate for the survey was 96%.

7. Ratings of local guides or lecturers at specific locations are not included. Most groups had only one guide, but some had a second guide join the group for all or part of the tour. When there was more than one guide, guides either divided the guiding of sites between them (i.e., each leading at a different site) or shared responsibility of guiding (i.e., both contributed to the guiding at the site). Observers were instructed to rate each guide separately, both on items relating to quality of guiding and on items relating to group management. Ratings for guides used in the analysis reflect the average rating for all regular guides on the trip.

8. The rating scales used by the observers were themselves composed of indicators to guide the observers in their ratings. For example, see the following scale for the guide's command of English:

#### The guide's command of language (fluency)

---

1. Poor command of the language. Needs assistance of a translator.
  2. ↑↓
  3. Working knowledge of the language. Some difficulty communicating effectively.
  4. ↑↓
  5. Excellent command of language. Fluent or native speaker.
- 

9. Reports from one bus on site presentations were not available.

10. Typically, correlations of .3 are significant in survey data because the number of cases tends to be over one thousand. Given that surveys are blunderbuss instruments, analysts are satisfied with statistically significant results and do not expect a single correlation to account for much variance.

11. All names are pseudonyms. The Hebrew pseudonyms we chose are the names of various trees.

12. All names of the guides are pseudonyms.

13. Charismatic guiding is likely a less important factor in trip success than overall group management skills—a further hypothesis for which our data provide the following suggestive evidence: The correlation between the factor in the observer ratings for "general effectiveness of guide" (a measure of group management) and the participants' questionnaire factor for "learning about Israel and Judaism" is .45 (significant at .05). The correlation between the observer factor for "site performance" (a measure of the guide's charisma) and the questionnaire factor for learning is not statistically significant. See Table 3.

14. Observations were approximately a week apart.

## REFERENCES

- Bond, T. G., and M. Fox. 2001. *Applying the Rasch model: Fundamental measurement in the human sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Creswell, J. W., V. L. Plano Clark, M. L. Gutmann, and W. E. Hanson. 2003. Advanced mixed methods research designs. In *Handbook of mixed methods in social & behavioral research*, edited by A. Tashakkori and C. Teddlie, 209–40. Thousand Oaks, CA: Sage.
- Goldberg, H., S. Heilman, and B. Kirshenblatt-Gimblett. 2002. *The Israel experience: Studies in youth travel and Jewish identity*. Jerusalem, Israel: Studio Kavgraph, Andrea and Charles Bronfman Philanthropies.
- Goldwater, C. 2002. Constructing the narrative of authenticity: Tour educators at work in the Israel experience. Master's thesis, Hebrew University of Jerusalem, Israel.
- Hanson, W. E., J. W. Creswell, V. L. Plano Clark, K. S. Petska, and J. D. Creswell. 2005. Mixed methods research designs in counseling psychology. *Journal of Counseling Psychology* 52 (2): 224–35.
- Henry, N. W., and P. F. Lazarsfeld. 1968. *Latent structure analysis*. Boston: Houghton Mifflin.
- Ivankova, N. V., J. W. Creswell, and S. L. Stick. 2006. Using mixed-methods sequential explanatory design: From theory to practice. *Field Methods* 18 (1): 3–20.
- Johnson, R. B., and A. J. Onwegbuzie. 2004. Mixed methods research: A research paradigm whose time has come. *Educational Research* 33 (7): 14–26.
- Kelner, S. 2002. Almost pilgrims: Authenticity, identity and the extra-ordinary on a Jewish tour of Israel. PhD diss., Department of Sociology, City University of New York.
- Miller, S. I., and J. L. Gatta. 2006. The use of mixed methods models and designs in the human sciences: Problems and prospects. *Quality & Quantity* 40 (4): 595–610.
- Perneger, T. V. 1998. What's wrong with Bonferroni adjustments? *The British Medical Journal* 316 (7139): 1236–38.
- Sale, J. E. M., L. H. Lohfeld, and K. Brazil. 2002. Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality & Quantity* 36:43–53.
- Saxe, L., C. Kadushin, S. Kelner, M. I. Rosen, and E. Yereslove. 2002. *A mega-experiment in Jewish education: The impact of birthright israel*. Waltham, MA: Cohen Center for Modern Jewish Studies, Brandeis University.
- Saxe, L., C. Kadushin, H. Shahar, and M. I. Rosen. 2004. *birthright israel: Recent findings and long-term effects*. Waltham, MA: Cohen Center for Modern Jewish Studies, Brandeis University.
- Saxe, L., T. Sasson, and S. Hecht. 2006. *Taglit-birthright israel: Impact on Jewish identity, peoplehood, and connection to Israel*. Waltham, MA: Cohen Center for Modern Jewish Studies, Brandeis University.
- Sieber, S. D. 1973. The integration of fieldwork and survey methods. *American Journal of Sociology* 78 (6): 1335–59.
- Tashakkori, A., and C. Teddlie. 2003. *Handbook of mixed methods in social & behavioral research*. Thousand Oaks, CA: Sage.

*CHARLES KADUSHIN is a Distinguished Scholar at the Cohen Center for Modern Jewish Studies, Brandeis University, and a visiting research professor in sociology at Brandeis. His research interests are in survey methodology, social networks, and Jewish studies. Three recent publications are The American Intellectual Elite (Transaction Publishers, 2006 [republishing of the 1974 edition with a new introduction]); "Personal*

*Influence: A Radical Theory of Action*" (The Annals of the American Academy of Political and Social Science, 2006); and (with Laurence Kotler-Berkowitz) "Informal Social Networks and Formal Organizational Memberships among American Jews: Findings from the National Jewish Population Survey 2000–01" (Sociology of Religion, 2006).

SHAHAR HECHT is a research associate at the Cohen Center for Modern Jewish Studies at Brandeis University. Research interests include Jewish young adults and Israel-Diaspora relations. Two recent publications are (with L. Saxe and T. Sasson) *Israel at War: The Impact of Peer-Oriented Israeli Programs on Responses of American Young Adults* (Cohen Center for Modern Jewish Studies, 2006) and (with L. Saxe and T. Sasson) *Taglit-birthright israel: Impact on Jewish Identity, Peoplehood, and Connection to Israel* (Cohen Center for Modern Jewish Studies, 2006).

THEODORE SASSON has a BA from Brandeis University and a PhD from Boston College. He is an associate professor in the Sociology-Anthropology Department, Middlebury College, and a Senior Scholar at the Cohen Center for Modern Jewish Studies, Brandeis University. He is coauthor (with Ephraim Tabory) of "A House Divided: Grassroots National Religious Perspectives on the Gaza Disengagement and Future of the West Bank" (Journal of Church and State, 2007); coauthor (with Shaul Kelner) of "From Shrine to Forum: Masada and the Politics of Jewish Extremism" (Israel Studies, forthcoming); and coauthor (with Katherine Beckett) of *The Politics of Injustice: Crime and Punishment in America, 2nd edition* (Sage Publications, 2004).

LEONARD SAXE (PhD) is a professor of Jewish community research and social policy at Brandeis University, where he also serves as director of the Cohen Center for Modern Jewish Studies and the Steinhardt Social Research Institute. Professor Saxe is a social psychologist and methodologist. His current research focuses on socio-demographic studies of the Jewish community, evaluative research of Jewish education programs, and policy research on a broad range of social issues. He coauthored a 2004 book about Jewish summer camping, *How Goodly Are Thy Tents* (University Press of New England), and is the author of a forthcoming book about birthright israel.