Department of Defense
Legacy Resource Management Program

PROJECT 11-157

An Experimental Test of the Accuracy and Adequacy of In-Field Artifact Analysis

Michael Heilen, with contributions by Jeffrey H. Altschul, Bradley J. Vierra, Rein Vanderpot, and Robert A. Heckman

March 2013

An Experimental Test of the Accuracy and Adequacy of In-Field Artifact Analysis

by Michael Heilen

with contributions by Jeffrey H. Altschul, Bradley J. Vierra, Rein Vanderpot, and Robert A. Heckman

Prepared with funding from the Department of Defense Legacy Resource Management Program Project No. 11-157

Prepared under contract with U.S. Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory
Contract Number W9132T-11-2-0014

Sponsored by the U.S. Army Garrison, Fort Huachuca, Arizona

Supported by Fort Bliss Military Reservation

Technical Report 12-90
Statistical Research, Inc.
Tucson, Arizona

Fort Huachuca Cultural Resources Report FH-12-20

March 2013

CONTENTS

LIST OF FIGURES

LIST OF TABLES

EXECUTIVE SUMMARY

In the western United States, the Department of Defense (DoD) has in some cases depended on in-field analysis for analyzing most or all artifacts used to characterize and interpret archaeological sites recorded during survey. Because the funding and space needed for artifact curation are sparse, the DoD and other federal agencies have justified in-field analysis with the need to limit curation. DoD policies that promote in-field artifact analysis over laboratory analysis rely on the assumptions that in-field analysis is of similar quality to laboratory analysis and that future research and management efforts will not require access to artifacts. Despite increasing reliance on in-field analysis for inventory and evaluation, the validity of these assumptions has not been tested. Legacy Project No. 11-157 was designed as a preliminary experiment to test those assumptions by assessing the adequacy and accuracy of in-field artifact analysis and digital-photograph analysis at two prehistoric archaeological sites located on military installations in the western United States: the Soldier Creek site (AZ EE:7:164 [AMS]), in southeastern Arizona, at Fort Huachuca, and FB 9583, in south-central New Mexico, on the East McGregor Range of Fort Bliss Military Reservation.

At each of these sites, samples of individually numbered ceramic and lithic artifacts discovered in surface contexts were analyzed by two separate field technicians. The artifacts were then collected and analyzed by trained specialists who analyzed, in a laboratory setting, either the physical artifacts or only digital photographs of the artifacts. The accuracy and precision of the in-field and digital-photograph analyses were assessed using several statistics: *agreement index, Cohen's kappa, McNemar's test for bias, false-positive rate*, and *false-negative rate*. For most assessments, the hands-on laboratory analysis was treated as the "gold standard" for the project, and the results of the in-field and digital-photograph analyses were tested against that standard. The adequacy of results for site interpretation and management was assessed by evaluating whether and how differences in analysis results could influence how sites are interpreted and managed.

In general, the results of these assessments showed that both the in-field and the digital-photograph analyses were of low accuracy and were often inadequate for site interpretation. Rare and important artifact types were often misclassified, and evidence for both random error and systematic bias in artifact identification was common. Digital-photograph analysis tended to be more precise than in-field analysis, but digital-photograph analysis also tended to identify rare types incorrectly, resulting in more-precise but inaccurate inferences about the temporal and cultural affiliations of a site. Several of the variables that many investigators have indicated as crucial to determining significance and establishing representative samples of sites for long-term preservation are derived primarily from artifact analysis: site function, assemblage diversity, temporal affiliation, and cultural affiliation. The experiment showed that all of those variables could be inaccurately assessed, sometimes grossly so, when based on in-field or digital-photograph analysis.

Following these results, a series of recommendations was made for deciding how and in which situations in-field analysis is best applied and regarding how to improve the reliability of in-field analysis. It was recommended that the decision to perform in-field analysis be made in consultation with stakeholders and based on the requirements and regulatory contexts of particular projects, not by blindly following no-collection or limited-collection policies. Methods and tools for improving quality and consistency in artifact identification should be developed through training, field guides, and implementation of digital technologies. The accuracy and adequacy of in-field-analysis data collected during previous projects at military installations should also be tested to identify situations in which data are likely to be faulty and how future data-collection efforts could be improved.

ACKNOWLEDGMENTS

The infield analysis project was designed by Jeffrey Altschul and Michael Heilen of Statistical Research, Inc. (SRI), and Martyn Tagg (Cultural Resources Manager, Fort Huachuca). Robert Heckman of SRI served as the project manager, and Michael Heilen served as the principal investigator. Teresita Majewski of SRI coordinated the scheduling and contracting of the project with contract administrators and with the Legacy Resource Management Program.

We would like to acknowledge the support of Martyn Tagg at Fort Huachuca as Department of Defense project sponsor. Brian Knight (Chief, Conservation Branch Environmental Division, Directorate of Public Works, Fort Bliss) facilitated Fort Bliss' contribution to the project and allowed SRI to incorporate important comparative data into the analyses. David Doyel (Archaeologist, 56 Range Management Office [56 RMO], Luke Air Force Base [AFB]) and Carol Heathington (Strategic Operations Planner, Luke AFB) also offered support, but were ultimately unable to fully participate in the project due to scheduling conflicts. We would like to thank Gregory E. Denman (Grants Officer, Engineer Research and Development Center, Construction Engineering Research Laboratory [ERDC-CERL]), and Anne M. Cox (Contract Specialist, ERDC-CERL) for administering the contract. Arthur Vokes (Curator, Arizona State Museum) provided access to previous collections from the Soldier Creek site on Fort Huachuca that were used during the project. Jonathan Gholson (GIS Analyst, Integrated Training Area Management, Fort Huachuca) created a location map of the Soldier Creek site for use during the project.

A number of individuals and organizations consulted on the project and provided guidance on consultation. Ann Howard (Deputy State Historic Preservation Officer [SHPO] for Arizona) reviewed the initial memorandum of Agreement (MOA) and work plan for work at Fort Huachuca and provided examples of MOAs and information on Arizona SHPO's views on artifact collection. Peter Steere, (Tribal Historic Preservation Officer [THPO], Tohono O'odham Nation), Angela Garcia-Lewis (Native American Graves Protection and Repatriation Act [NAGPRA] coordinator, Salt River Pima-Maricopa Indian Community), and Mark Altaha (THPO, White Mountain Apache Tribe) provided input that helped to steer the project in the right direction. Katharine Kerr (Program Analyst, Advisory Council on Historic Preservation) provided input on MOAs and on project methods in consideration of tribal concerns. Ms. Heathington at Luke AFB provided input on the consultation process.

Kelly Merrifield deserves thanks for graciously working with us to complete the project on schedule as she transitioned into her new role as Cultural Resources Management Specialist, Legacy Resource Management Program. We also want to thank Cecilia Brothers (former Cultural Resources Management Specialist, Legacy Resource Management Program) for her encouragement and patience and for sticking with the project even as she transitioned into a new position with the Center for Environmental Management of Military Lands. Sean Rutherford (Assistant Resource Management Specialist, Legacy Resource Management Program) also provided important information and assistance during critical stages of the project.

This project depended on the work of many individuals at SRI who performed fieldwork, laboratory analysis, project administration, and report production. Fieldwork was completed by David Unruh, William White, Brandon McIntosh, Amanda Hernandez, and Donovan Quam. Laboratory work was completed by Bradley Vierra, Rebecca Kiracofe, Monica Murrell, Meaghan Trowbridge, and Lauren Jelinek. Project administration was performed by Lisa Atkinson, Jody Holmes, Teresita Majewski, Rein Vanderpot, and Robert Wegener. Report production was directed by Maria Molina. Beth Bishop and Grant Klein edited the report; Linda Wooden and Jason Pitts formatted and laid out the report; and Andrew Saiz worked on the report figures.

Finally, we would like to thank the following individuals for providing guidance or comments on the draft report: Cecilia Brothers (now at Center for Environmental Management of Military Lands), Brian Knight, Kathleen McLaughlin (Army Deputy Federal Preservation Officer), Kelly Merrifield, and Martyn Tagg.

CHAPTER 1

Introduction

As part of a U.S. Department of Defense (DoD) Legacy Resource Management Program (Legacy) project (No. 11-157), this report explores the accuracy and adequacy of in-field artifact analysis. This project was sponsored by Fort Huachuca and was conducted for Legacy by Statistical Research, Inc. (SRI), through Contract No. W9132T-11-2-0014 with the U.S. Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory. Both Fort Huachuca and Fort Bliss were contributors to this project. The project consists of an experiment to test the accuracy and adequacy of in-field analysis conducted at two project sites—one on Fort Huachuca, in southeastern Arizona, and one on Fort Bliss Military Reservation, in south-central New Mexico (Figure 1). At each of these sites, a sample of ceramic and lithic artifacts was documented by two separate field crews, resulting in two separate sets of field identifications for the same, individually numbered artifacts. Digital photographs of the same artifacts were also taken in the field. The documented artifacts were then collected and subsequently analyzed in a laboratory setting. Laboratory analysis was conducted by trained specialists who performed analysis using either (1) the physical artifacts or (2) only digital photographs of the artifacts. This process resulted in a total of four sets of independently obtained analysis results for each of the two project sites.

In this report, the results of these analyses are assessed in order to test for the level of agreement between artifact identifications made in the field by archaeological field technicians and those made in the laboratory by trained specialists using digital photographs only or the physical artifacts. The following questions were addressed:

- What is the level of agreement between observers, according to observational method, training, artifact attribute, and site?
- How accurate or precise are artifact observations made in the field or using only digital photographs?
- To what degree could variation in the accuracy and precision of artifact identification affect the interpretation of a site, in terms of site function, technology, temporal affiliation, or cultural affiliation?

Currently, scientific studies that have tested the accuracy and adequacy of in-field analysis at DoD installations are rare. Inaccurate identification of artifacts during in-field analysis can result from a variety of factors, including lack of adequate lighting and appropriate analytical instruments, unclean surfaces, adverse weather conditions, and experience of field personnel. The limited research that has been conducted on this subject is not encouraging. Two pilot studies indicated that the problem is worthy of serious attention from the DoD. First, blind studies performed at the Valencia site in the Tucson Basin, demonstrated that in-field analysts varied so widely on simple counts of plain ware body sherds that the results were largely meaningless (Altschul 1986). More recently, a Legacy-funded study of archaeological-data quality (Legacy Project No. 07-353) (Heilen et al. 2008), demonstrated that even for basic parameters, such as site location and site boundaries, the accuracy and reliability of field observations vary tremendously within and between DoD installations, regardless of Military Service or environmental conditions. The current project is a follow-on project to the Legacy-funded study of archaeological-data quality.

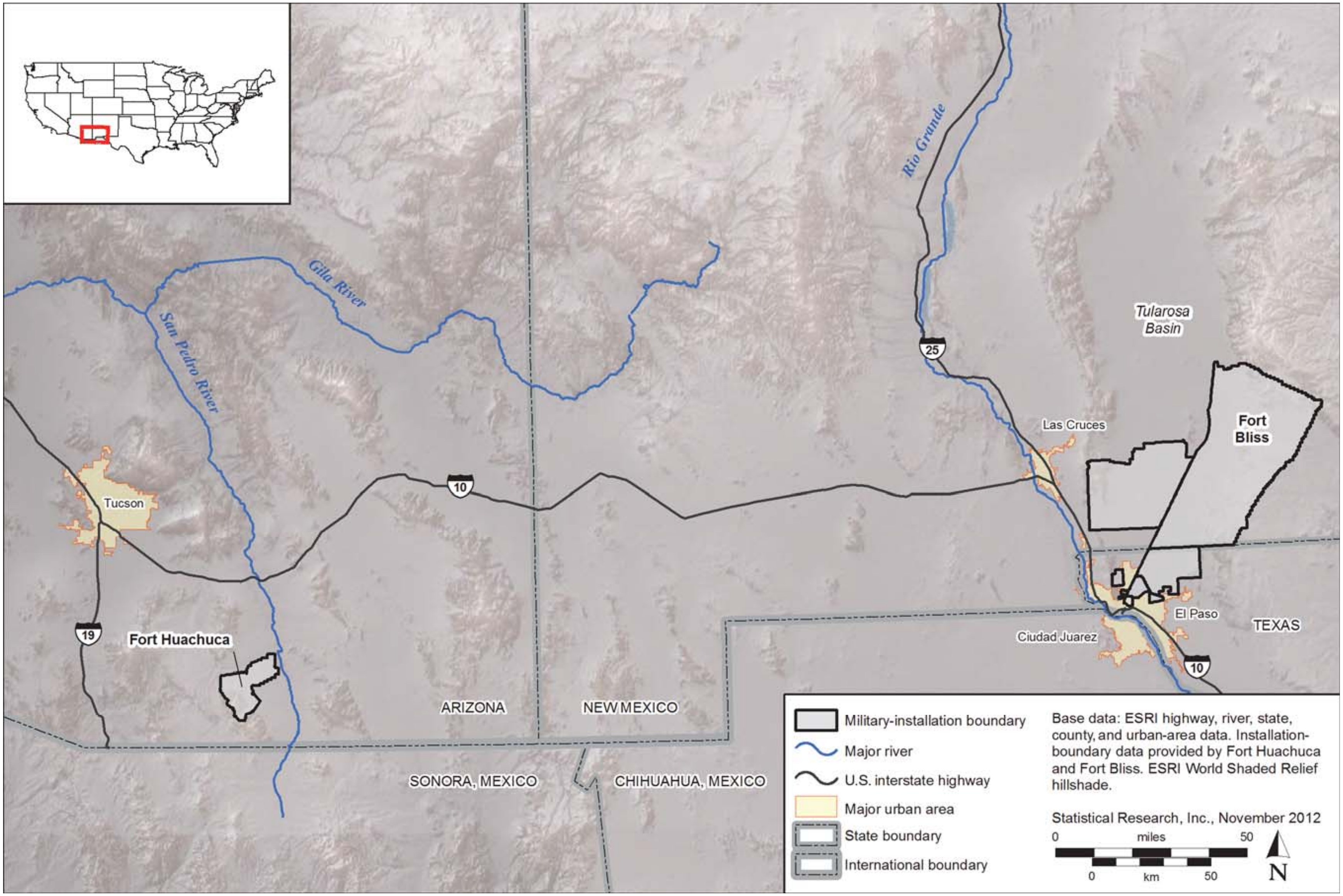

Figure 1. The locations of the two project sites within the states of Arizona and New Mexico.

Report Organization

This chapter introduces the project and provides background information regarding the purposes of current collection policies and their scientific and legal implications. The remainder of this report provides context for the two project sites; describes field, laboratory, and analytical methods; and evaluates, using multiple assessment methods, the results of the in-field artifact analysis and the digital-photograph and hands-on laboratory analyses. Chapter 2 provides a brief culture history relevant to the project sites and collected materials, describes the project sites, and presents the methods used for artifact analysis and for assessing analysis results. Chapter 2 also includes a description of the statistics calculated to assess the level of agreement among analyses as well as to assess the accuracy and precision of artifact identifications made by different observers. These include calculation of an agreement index, classification-error estimates, Cohen's kappa agreement statistic, and McNemar's test for bias. The distribution of results among analyses is also compared to determine where and in what cases substantial differences in site interpretation could have resulted from differences in analysis results. Chapter 3 uses these methods to assess the analysis results for ceramic artifacts collected from the two project sites. Chapter 4 assesses the analysis results for lithic artifacts collected from the two project sites. The final chapter of this report, Chapter 5, provides a summary of the project approach, methods, and results and offers recommendations for best practices in regard to in-field analysis and artifact collection.

The Limiting of Collection during Survey

It is longstanding archaeological practice that artifacts collected in the field must be documented and curated at a museum or repository. This practice has served archaeology well. Insights are continually gleaned from old collections. Artifacts of rare and exotic types, often targeted by collectors and vandals, now exist almost exclusively in repositories. Without these collections, archaeologists, indigenous communities, and the general public would lose a valuable connection with the past and their heritage.

For most of the twentieth century, archaeologists made collections from sites during survey. Sometimes, these collections were systematic, following either a statistically based sampling scheme or a procedure aimed at ensuring that collections at sites were made consistently. More often, collections were grab-bag samples, during which archaeologists made subjective appraisals of the artifacts on the surface and collected samples that reflected their interests. In general, temporally and/or culturally diagnostic materials were the primary targets of these collections.

In many parts of the western United States, the practice of collecting artifacts during survey began to be challenged in the 1980s. The Arizona State Historic Preservation Office (SHPO) took a particularly aggressive stance, arguing that the collection of artifacts could be considered an adverse effect. For surveys conducted in compliance with Section 106 or 110 of the National Historic Preservation Act (NHPA), the sponsoring government agencies were required to consult with SHPOs and the affiliated Native American tribes and communities about their survey plans. Survey plans that included collections received more attention and required agencies to justify why collections were necessary, how the collections were going to be analyzed, and where they were going to be curated. Why, in essence, couldn't the artifacts be identified in the field and left there?

Some agency archaeologists took on the added burden of justifying collections, but many did not, not simply because of the extra work but also because government agencies had a vested interest in limiting collections. By eliminating collections, agencies decreased their costs significantly; they did not have to analyze artifacts in the laboratory, prepare the artifacts for curation, and pay for a collection to be curated in perpetuity. Not only government-agency and SHPO archaeologists favored the trend toward limited collection; without having to process, analyze, and curate collections, the reduced level of effort associated with no-collection and limited-collection policies also allowed contractors to reduce project costs or to put more effort into other project tasks.

By the end of the 1980s, the practice of systematic artifact collection during surveys had been all but eliminated in Arizona. Collection was limited to artifacts of rare types or types favored by vandals and collectors (e.g., projectile points, whole ceramic vessels, and historical-period bottles). Arizona was by no means alone in the trend to eliminate or limit surface collection during survey. California, New Mexico, Utah, and Nevada were among the states that followed similar tactics.

Implicit in no-collection and limited-collection policies is the assumption that the observations made about artifacts in the field are as accurate as those made in the laboratory. Surprisingly, in the past 30 years, there have been very few studies performed that have actually tested this assumption. Altschul (1986) examined data from the Valencia site in Tucson and showed convincingly that there were wildly different results between data collected from transects in which artifacts simply were counted and data collected from transects from which artifacts were intensively collected. This cautionary case study, however, did not stem the tide toward adoption of a limited-collection policy in Arizona.

No-collection and limited-collection policies have not been without their critics. Commonly, evaluation and data recovery projects based on no-collection or limited-collection policies uncover very different archaeological resources from what has been anticipated or previously reported. Such results leave private and public sponsors of archaeological investigations, along with Native American tribes, wondering why archaeologists can so often be wrong. Some agencies, such as the Tonto National Forest, routinely revisit sites recorded by contractors, because they have little faith that artifact identification has been done correctly (Michael Sullivan, personal communication 2010). Others, such as the Coconino National Forest, require an artifact-identification test before permitting archaeologists to work within the forest. These examples suggest that the public may not have been well served by no-collection or limited-collection policies.

Evaluating no-collection or limited-collection policies requires baseline data that compare artifact identification in the field with artifact identification conducted in the laboratory. This study is a first step in that direction. But even if we find that no-collection or limited-collection policies result in inadequate artifact identification which then leads to inaccurate resource assessments that can negatively affect management decisions, it is unlikely that we can return to the days of systematic surface-artifact collection. Like many federal agencies in the western United States, the DoD has partially embraced no-collection or limited-collection policies for situations in which it makes sense to apply them. This is in large part because the DoD recognizes that it has increasingly limited space in which to curate artifacts and that additional funding is needed to support curation. It is important to note that the Military Services each manage their respective lands under the auspices of their individual programming and policies. Generally, DoD installations in the western United States selectively follow a no-collection or limited-collection policy, as appropriate, based on whether they have discretion over who will actually be conducting the fieldwork. **The goal of this project, then, is not simply to compare in-field artifact analyses with other procedures but to identify ways to make no-collection or limited-collection policies work more efficiently, accurately, and effectively.**

To understand the manner in which we approached the problem, we need to place the no-collection and limited-collection policies used by the DoD in context. The remainder of this introduction discusses curation from a federal-agency perspective. We begin with a federal agency's curation responsibilities. We then describe the evolution of no-collection and limited-collection policies and their critics, especially as these relate to the DoD. Lastly, we describe the need for this project and how it is organized and presented.

Federal Curation Policy

Federal legislation and regulations mandate that collections resulting from archaeological projects conducted on federal property or benefiting from federal funding be maintained and curated in perpetuity.

Collections include artifacts and materials collected during the course of an archaeological investigation as well as any associated records. Curation refers to the care, management, and use of collections. Per professional ethical standards and federal guidelines, properly curated collections must be preserved to prevent deterioration; must be appropriately documented and organized; and must be made accessible for scientific research, public education, and cultural use. Procedures and guidelines for managing collections recovered by federal projects have been established in Title 36, Part 79, of the *Code of Federal Regulations* (36 CFR 79): Curation of Federally-Owned and Administered Archaeological Collections. Pursuant to Section 101(a)(7)(A) of the NHPA (U.S. Code Title 16, Section 470*a* [16 U.S.C. 470*a*]), 36 CFR 79 establishes "definitions, standards, procedures and guidelines to be followed by Federal agencies to preserve collections of prehistoric and historic material remains, and associated records, recovered under the authority of the Antiquities Act (16 U.S.C. 431–433), the Reservoir Salvage Act (16 U.S.C. 469–469*c*), Section 110 of the National Historic Preservation Act (16 U.S.C. 470*h*-2) or the Archaeological Resources Protection Act (16 U.S.C. 470*aa–mm*)." Two previous Legacy projects (*Guidelines for the Field Collection of Archaeological Materials and Standard Operating Procedures for Curating Department of Defense Archaeological Collections* [Legacy Project No. 98-1714] and *Archaeological Collections Management Procedures* [Legacy Project No 06-319]) have provided guidance for DoD installations regarding collection procedures and collections management and curation (see Griset and Kodack 1999; Sagebiel et al. 2007).

These and other studies have shown that although collection and curation preserve the scientific basis of archaeology and support the effective stewardship of heritage resources, collection and curation do not come without a cost. Over the past several decades, a curation crisis has been mounting in cultural resource management (CRM) in the United States. As development increases, so do CRM activities. The volume of collected materials requiring curation has increased in tandem with development, whereas the space available for curation has remained static, funding for curation has decreased, and conditions for ensuring the preservation of collections are inadequate. Moreover, many materials that were collected during early investigations, along with the associated documents, have been lost or destroyed or are stored in conditions that are inadequate for ensuring their preservation. In many states, the collections that have been made and where they are held have been poorly documented. Efforts to develop centralized databases documenting the locations and statuses of collections have only just begun. The result has been a curatorial problem of crisis proportions, whereby collections of vast numbers of irreplaceable artifacts of scientific and cultural value have not been adequately curated and are stored in conditions not conducive to the long-term preservation of archaeological materials and associated records (Bawaya 2007; Lyons et al. 2006).

A number of professional associations and governmental entities have called attention to this crisis (Heyman 1997; Lyons et al. 2006; Nepstad-Thornberry et al. 2002; Sullivan 1992; Trimble and Meyers 1991), including the U.S. Army Corps of Engineers (USACE), the U.S. Department of the Interior (USDI) National Park Service (NPS), the Smithsonian Institution, the Society for American Archaeology (SAA), the Society for Historical Archaeology, the Arizona Governor's Archaeology Advisory Commission Curation Subcommittee, the Executive Committee of the Colorado Council of Professional Archaeologists, and the Council of Virginia Archaeologists' Collections Committee, among others. Recommendations offered by these entities for remedying the crisis have addressed the following concerns:

- recognizing and documenting the extent of the curation crisis;
- underscoring the scientific, cultural, and legal imperatives for curation;
- garnering public support;
- developing and lobbying for funding mechanisms to address the problem and improve curation outcomes;
- increasing curation space and improving curatorial procedures; and
- developing strategies for limiting the volume of collections.

Other specific recommendations for addressing the crisis have included the following:

- deaccessing redundant materials to regain curation space;
- obtaining federal support for the rehabilitation of existing federal collections;
- upgrading curation facilities that house federal collections;
- developing inventories of all federal collections held in repositories;
- developing an accreditation program for repositories;
- developing national standards and guidelines for making field collections and for accessing, using, and managing collections and associated records; and
- establishing a national archaeological-curation program that would include the creation of an Office of Archaeological Curation within the USDI and the addition of a curation position within the SHPO of each state.

Limiting Collections: One Means of Addressing the Curation Crisis

The above recommendations have mostly to do with developing funding, standards, and procedures for the management of existing collections or improving the conditions under which new collections are managed. Another seemingly practical, if only partial, solution to this crisis is either to limit collections or not to collect at all (Beck and Jones 1994; Butler 1979). The NPS has defined a collection as "material remains that are excavated or removed during a survey, excavation, or other study of a prehistoric or historic resource, and associated records that are prepared or assembled in connection with the survey, excavation or other study" (NPS 2011). In the context of limited-collection or no-collection policies, "collection" refers principally to material remains and not to associated records that document or record those remains. Representative sampling of archaeological materials from archaeological investigations for curation is one means of limiting collections and dampening the volumetric increase in collections that require curation. The approach assumes, of course, that archaeologists will be able to devise sampling strategies that adequately address legal requirements and scientific needs, now and in the future, and safeguard against the loss of important information. Not collecting artifacts or limiting their collection means that many or all artifacts associated with an archaeological site must be analyzed in the field. In-field analysis is often done by field crew with limited training in artifact analysis, rather than by trained specialists, and it is often done under conditions that may be adverse to achieving accurate results, such as inclement weather, unavailability of adequate measurement tools, poor lighting, or unclean surfaces.

For more than 20 years, limited-collection and no-collection policies for archaeological surveys have been in effect throughout most of the western United States, including at most DoD installations. U.S. Air Force Instruction 32-7065, for instance, calls for the establishment of procedures in Integrated Cultural Resource Management Plans to minimize, during inventory or excavation, the numbers of collected materials that require permanent curation. Similarly, the U.S. Army (Army) discourages collection and directs installations to limit collections, particularly during survey, by performing in-field analysis and limiting the permanent curation of materials collected during mitigation to "diagnostic artifacts and other significant and environmentally sensitive material that will add important information to site interpretation" (Army 2007:Section 6-4.e[5]). Identification of surface artifacts is performed in the field by field crews, and only rare artifact types, such as projectile points and whole vessels, are collected.

As a result of limited-collection and no-collection policies, site classifications and, in many cases, National Register of Historic Places (NRHP) evaluations are based on the results of in-field analyses. The Arizona SHPO has also begun investigating the circumstances under which in-field analysis is conducted, such that it may be conducted not only during inventory or evaluation projects but also during data recovery projects. Other states are poised to follow Arizona's example.

The Impact and Implications of Limited-Collection and No-Collection Policies

A recent Legacy-funded project (*Archaeological Collections Management Procedures* [Legacy Project No. 06-319]) showed that within the DoD, collections generally are not made during surveys, although artifacts are often described, measured, photographed, and drawn in the field (Sagebiel et al. 2007). In some cases, collections are restricted to diagnostic artifacts (those that provide cultural or temporal information) or artifacts that are in imminent danger of destruction or of removal through illegal collection activities. If shovel tests are necessary during survey because of ground cover, then any artifacts uncovered are collected. The particular method and sampling strategy for collections made during testing and data recovery excavations are not usually covered in federal-agency guidelines. Instead, those decisions are left to lower-level entities (military installations, individual forests, etc.) or are determined on a project-by-project basis. Usually, federal entities follow the local-SHPO guidelines regarding collection.

In 2003, the SAA Advisory Committee on Curation noted that the development or implementation of no-collection policies by federal and state agencies had been haphazard, at best. The committee further observed that limiting collections during fieldwork "has huge implications for the future quality and usability of the collections cared for in the public and professional interest. Standards are needed to help determine the kinds and types of artifacts to be collected during survey, site testing, and excavation. Guidance is also needed on developing statistically valid sampling strategies and documenting all decision-making" (SAA Advisory Committee on Curation 2003:4).

In their report, *Guidelines for the Field Collection of Archaeological Materials and Standard Operating Procedures for Curating Department of Defense Archaeological Collections* (Legacy Project No. 98-1714), Griset and Kodack (1999:32) stated four reasons that no-collection surveys should be avoided:

1. Field identification of artifacts cannot be verified without a costly return to the field.
2. Future investigations, such as archaeometric tests, are not possible without a costly return to the field to collect a sample.
3. No-collection policies are rarely strictly adhered to, leaving a biased and skewed sample.
4. A no-collection policy only works to conserve the archaeological record if it is followed up by intensive and aggressive monitoring practices, which are also costly.

The potential problems with no-collection policies are only amplified if applied to testing and data recovery efforts, because there may be no opportunity to verify results or to apply new artifact-analysis techniques. Instead of no-collection policies, Griset and Kodack (1999:32) recommended collection based on practical, scientific, and replicable principles of statistical sampling.

The decision not to curate artifacts marks a break with archaeological tradition dating back to the nineteenth century and places a tremendous burden on field archaeologists to make sure that their artifact identifications are correct. But how accurate are in-field analyses? What classes of artifacts or artifact attributes can be recorded adequately in the field? Which artifact classes need laboratory analysis? Also, which are likely to be revisited by future researchers in the laboratory? The answer to these questions is "No one knows." The purpose of this project is to begin to provide the DoD with good answers upon which decisions about field recording and future curation can be based.

CHAPTER 2

Context and Methods

As discussed in Chapter 1, this project is an experimental project aimed at assessing the adequacy and accuracy of in-field and digital-photograph analyses of ceramic and lithic artifacts in comparison to hands-on laboratory analysis of the same artifacts. In order to replicate standard conditions under which CRM investigations are conducted in the western United States, the in-field analysis was performed by field technicians, and the digital-photograph and hands-on laboratory analyses were conducted by specialists with specific training in ceramic or lithic analysis. Analyses were performed using a sample of artifacts collected from two sites. The results of the analyses were assessed using a variety of quantitative and qualitative methods for assessing the level of agreement between observers.

In this chapter, information is presented on the two sites investigated during the project, along with information on field, laboratory, and analysis methods. One of the sites is located in southeastern Arizona, on Fort Huachuca. The site is on a ridge that overlooks a floodplain of Soldier Creek, on the western side of the San Pedro River valley. The other site is in south-central New Mexico, on Fort Bliss Military Reservation. The site is located on the eastern side of the Tularosa Basin, on an alluvial terrace of El Paso Draw. Both of the sites investigated for the project are prehistoric habitation sites that date primarily to the Formative period. At each site, a limited sample of ceramic and lithic artifacts was analyzed and collected.

To place the two investigated sites in context, the chapter begins with a brief culture history pertinent to both sites. This is followed by specific information on each of the two sites, including information on field methods or conditions specific to either site. After the culture history and site information are presented, general methods and tools for analyzing the artifacts are presented. The chapter ends with a discussion of the analytical approach and statistical methods used to evaluate the consistency and reliability of the artifact-analysis results.

Culture History

This section presents the culture history pertinent to the interpretation of the two sites investigated during the project. The culture history draws heavily from Heilen et al. (2012), Miller et al. (2009), Vanderpot (2012), and Vierra (2012). The culture history focuses primarily on the prehistoric period in southeastern Arizona (where Fort Huachuca is located) and south-central New Mexico (where the site investigated on the Fort Bliss Military Reservation is located), because the artifacts analyzed as part of the project were derived from prehistoric components at the two sites. The culture history begins with the Paleoindian period and ends with the protohistoric and historical periods.

The Paleoindian Period

The earliest people of the Americas were highly mobile hunter-gatherers who depended on large mammals, including now-extinct megafauna, for their subsistence. The Paleoindian period is traditionally divided into three periods, based on technology and land use: the Clovis, Folsom, and Late Paleoindian periods

(Holliday 1997; Stanford 2005). In southeastern Arizona, the Paleoindian period is defined as dating from ca. 11,500 to 8500 B.C. A longer and somewhat different span is attributed to the Paleoindian period in south-central New Mexico, ca. 10,000–6000 B.C.

Of the several Paleoindian traditions, Clovis was the earliest and most widespread, having been found throughout much of North America (Haynes 1993). Located in the San Pedro River valley near Fort Huachuca is one of the most important concentrations of Clovis sites in North America; it includes the sites of Naco, Escapule, Leikem, Navarrete, and Murray Springs (Huckell and Haynes 2007:222). Paleoindian period sites are also likely present at Fort Huachuca, based on the discovery of a Clovis projectile point fragment, a probable Paleoindian point midsection (Tagg 2011), and several sites with Pleistocene faunal remains at Fort Huachuca (Cook 2003, 2004b, 2005b; Huckell 1982; Vanderpot 1997; Van West et al. 1997:146).

Paleoindian period sites discovered in south-central New Mexico have included those dating to the Clovis, Folsom, and Late Paleoindian periods, and Folsom occupations appear to be most common. Clovis finds in the vicinity of Fort Bliss have been comparatively few and scattered. During the Folsom period, the ancient environment of the Tularosa Basin was characterized by a patchy distribution of grasslands and playas. Folsom groups are believed to have followed small herds of bison moving between these foraging areas and to have established residential camps along the way. Most Folsom points were made on local materials, suggesting that the territories of these groups may have been restricted in size.

Late Paleoindian period groups appear to have focused their activities more on resources located in playa and ephemeral-stream settings than did their Folsom period predecessors. Late Paleoindian period sites in the Tularosa Basin have been found across a wide range of topographic zones, including mountains and alluvial fans. Most finds have been located near major playas or along the margins of the Rio Grande Valley (Carmichael 1986; Vierra 2012).

The Archaic Period

Following the Paleoindian period, the Archaic period was characterized by the broad-spectrum foraging of diverse plant and animal resources, including both large- and small-game animals and many different plant species. Increased focus on the collection and processing of plant materials, such as small seeds from annual plants, is evident during the period. By the end of the Archaic period, subsistence increasingly focused on the encouragement, protection, and cultivation of wild plants and cultigens (Doolittle and Mabry 2006).

The Archaic period in the Southwest was first defined in southeastern Arizona by Sayles and Antevs (1941), who defined the Cochise tradition as consisting of three phases: Sulphur Springs, Chiricahua, and San Pedro. Today, the Archaic period is conventionally divided into the Early, Middle, and Late Archaic periods. The precise boundaries of the Archaic period and its subdivisions differ between southeastern Arizona and south-central New Mexico. In southeastern Arizona, the Archaic period dates from 8500 B.C. to A.D. 1; the period is divided into the Early Archaic (8500–4800 B.C.), Middle Archaic (4800–1500 B.C.), and Late Archaic (1500 B.C.–A.D. 1) periods (Huckell 1984:136–142, 1995:16). In south-central New Mexico, the period dates from 6000 B.C to A.D. 200. There, the Archaic period is also divided into three periods, but the beginning and end dates are, compared to the periods defined for southeastern Arizona, shifted forward in time: Early Archaic (6000–4000 B.C.), Middle Archaic (4000–1200 B.C.), and Late Archaic (1200 B.C.–A.D. 200) (Miller and Kenmotsu 2004).

In both southeastern Arizona and south-central New Mexico, the Early Archaic period is poorly known, largely because of a scarcity of sites dating to the period. Most finds consist of isolated projectile points. The Early Archaic period sites that have been studied are characterized by milling equipment and flaked stone tools. The Middle Archaic period is somewhat better understood and is characterized by base camps that were positioned on the landscape to facilitate access to resources located in multiple environmental zones. Foraging strategies involved increased use of upland resources, including exploitation of pinyon nuts. Middle Archaic period sites were frequently located close to water sources (e.g., river valleys, springs, and playas), and groups staged logistical trips from such sites to obtain resources (Vanderpot 1997; Vierra 2007).

The Late Archaic period is characterized by the appearance of maize agriculture by at least ca. 1500 B.C. in southeastern Arizona and by 1200 B.C. in south-central New Mexico. Late Archaic period lifestyles focused on a mixture of foraging and farming subsistence practices, increased dependence on storage, decreased residential mobility and increased logistical mobility, and experimentation with ceramic technology. Reliance on maize agriculture may have been relatively low during that time, despite the appearance of cultigens in the region. Settlement during the period appears to conform to a bimodal settlement pattern in which lower-valley and alluvial settings were occupied in order to cultivate plants and exploit such resources as rabbits, mesquite, and seeds from annual plants, and upland environments were occupied in order to hunt artiodactyls and collect pinyon nuts and other plant foods. Some horticulture was performed at upland sites, as well (Anderson 1993; Bohrer 2007; Diehl 2005; Hard et al. 1996; Lentz 2006; Mabry 1998; MacNeish 1993; Roney and Hard 2002, 2005; Tagg 1996; Upham et al. 1987; Wegener et al. 2006; Whittlesey et al. 2007).

The Formative Period

The Formative period began with the widespread use of fully developed ceramic-container technology, a change that may relate to increasing reliance on cultigens or the storage and cooking of specific cultivars, such as beans and new varieties of floury maize (Deaver and Ciolek-Torrello 1995; Mabry 2005). Many important changes in settlement, land use, subsistence, and regional interactions occurred during the Formative period, resulting in a dynamic and varied archaeological record. Major changes, for instance, came with the appearance in Arizona, during the Middle Formative period, of the Hohokam, who built some of the largest irrigation systems in the Americas, and with their subsequent apparent disappearance during the Late Formative period (Whittlesey et al. 1994). A distinctive Mogollon culture also appeared during that time in southeastern Arizona, southern New Mexico, and northern Mexico.

In southeastern Arizona, the Formative period (A.D. 1–1450) is divided into the Early Formative (A.D. 1–700), Middle Formative (A.D. 700–1150), and Late Formative (A.D. 1150–1450) periods. The Middle and Late Formative periods, respectively, have been further subdivided into the Middle Formative A (A.D. 750–950) and Middle Formative B (A.D. 950–1150) periods and Late Formative A (A.D. 1150–1300) and Late Formative B (A.D. 1300–1450) periods. In south-central New Mexico, the Formative period (A.D. 200–1450) has been divided into the Early Formative (A.D. 200–1000), Middle Formative (A.D. 1000–1275/1300), and Late Formative (A.D. 1275/1300–1450) periods. These periods are also referred to as the Mesilla, Doña Ana, and El Paso phases, respectively.

The Early Formative Period

In both southeastern Arizona and south-central New Mexico, the Early Formative period corresponds to a transition from Archaic period lifestyles characterized by greater residential mobility and reliance on hunting and gathering to more-settled lifestyles characterized by an increasing reliance on agriculture, storage, village life, and ceramic-container technology. Agriculture appears to have become a central subsistence focus earlier in southeastern Arizona.

In southeastern Arizona, the first pottery to appear was a sand-tempered plain brown ware, in the form of either small, neckless jars or outcurved bowls (Whittlesey and Heckman 2000:6). Slipped and polished red wares characterized by larger vessel sizes and new forms appeared between A.D. 300 and 500. Brown pottery painted in simple red, broad-line patterns appeared throughout the region by A.D. 650 (Dean 1991). Subsistence relied partly on the farming of maize and other cultigens, and there was a corresponding decrease in residential mobility. Flaked stone technology during the period became increasingly expedient, involving a simple core-and-flake technology, although the biface and ground-stone-tool technologies that characterized the Late Archaic period continued to be used. During the Early Formative period in southeastern Arizona, people lived in bean-shaped structures and large, communal houses. Burials took the form of both semiflexed inhumations and cremations (Deaver and Ciolek-Torrello 1995:484–485). However, there is little direct evidence of Early Formative period occupations at Fort Huachuca.

In south-central New Mexico, site numbers and artifact densities associated with the Early Formative period increased dramatically from the preceding Archaic period. As in southeastern Arizona, the use of ceramic containers appeared during the Early Formative period in south-central New Mexico. Subsistence during the Early Formative period appears to have remained focused more on foraging than farming, however (Hard et al. 1996; Miller and Kenmotsu 2004; Whalen 1977, 1978). Early Formative period settlements were smaller than those of later periods and likely consisted of one household to a few households. Studies of macrobotanical remains have indicated that maize ubiquity was low during the Early Formative period and was substantially less than that reported from Late Formative period sites; reliance on cultigens may have been minimal. At present, it appears that cultivation supplemented a diet based on foraging for wild-plant foods (Adams 2009; Goldborer 1985; Miller and Burt 2007; Miller and Kenmotsu 2004). Interestingly, evidence for the use of the bow and arrow in south-central New Mexico is from late in the period; spears tipped with dart points continued to be used alongside the bow and arrow during that time, suggesting the continued use of weapons systems common to earlier periods (Kelley 1984; Miller and Kenmotsu 2004).

The Middle Formative Period

Clear differences in painted ceramic styles were apparent in southeastern Arizona by roughly A.D. 700, and these, combined with other salient changes in material culture and social organization, mark the beginning of the Middle Formative period there. Unique configurations of architecture, mortuary practices, iconography, and other aspects of material culture and site structure appeared during that time and signal the emergence of distinctive cultural traditions in southern Arizona, including the Hohokam, Mogollon, and Trincheras cultural traditions. The differences among these and other cultures in the greater Southwest increased throughout the period, such that different regions are often identified with specific groups.

For instance, in southeastern Arizona, distinctive Hohokam cultural traits, including buff ware pottery and Hohokam iconography, are concentrated in the Santa Cruz River valley. By the end of the period, Hohokam groups had developed large communities based on floodwater and irrigation farming, produced red-on-brown painted pottery, established a regional exchange and communication network associated with a ball-court complex, created a well-developed cremation burial ritual, and forged trade relations with cultures to the south, including Trincheras groups in present-day northern Mexico. Farther to the north, in the Phoenix Basin, the Gila-Salt Hohokam practiced similar lifeways but developed distinctive cultural traditions that included the production of red-on-buff painted pottery and the development of exceptionally large and sophisticated canal irrigation systems (Whittlesey et al. 1994).

Cultural traits associated with the San Simon branch of the Mogollon culture appeared in southeastern Arizona during that time (Sayles 1945; Van West et al. 1997; Whittlesey et al. 1994). During the early and middle part of the period, Mogollon groups produced polished brown plain ware pottery, red-slipped pottery, and red-on-brown painted pottery and constructed deep, bean-shaped or rectangular pit houses that were clustered in hamlets or small villages. Corrugated utilitarian pottery was added to the repertoire toward the end of the period, and surface structures built in contiguous room blocks replaced the pit-house villages. Room blocks contained rectangular kivas or ceremonial rooms, and they were built around public spaces.

On Fort Huachuca, Middle Formative period sites include habitations, rock-feature sites, and artifact scatters. Architectural features dating to the period appear to have been pit structures rather than surface structures. Most Middle Formative period sites on Fort Huachuca are clustered in two areas: along Soldier Creek and at the mouth of Garden Canyon. The Soldier Creek (AZ EE:7:164 [AMS]) and Garden Canyon (AZ EE:11:13 [ASM]) sites are the most important habitation sites dating to the Middle Formative period; both are interpreted as having been hamlets. Much of the pottery from these sites is similar to that of the Tucson Basin Hohokam, but some ceramic artifacts are similar in design to San Simon–series and

Dragoon-tradition pottery. Sherds corresponding to San Simon series or Dragoon tradition decorated types include Encinas Red-on-brown, Cascabel Red-on-brown, and Dragoon Red-on-brown. Trincheras Purple-on-red, San Francisco Red, and obliterated indented corrugated sherds also have been found, in rare instances, at Middle Formative period sites on Fort Huachuca.

In south-central New Mexico, the Middle Formative period (or Doña Ana phase) is considered to represent the pit-house-to-pueblo transition (Carmichael 1986). During that period, occupational intensity is inferred to have increased dramatically from the preceding Early Formative period, though not necessarily as a result of agricultural intensification. Middle Formative period sites in the southern Tularosa Basin are generally large, and they contain middens more frequently than do earlier and later sites. The frequent presence of pestles at Middle Formative period sites suggests the continued use of mesquite as an important resource; the abundance of rabbit bone in middens testifies to a focus on the hunting of small mammals.

Middle Formative period sites in the Tularosa Basin are typically identified through the detailed study of ceramic types and frequencies (Hard et al. 1994) but can be difficult to distinguish from multicomponent sites representing discrete occupations during multiple phases. In general, the Doña Ana phase is considered to represent increased reliance on agriculture and intensified plant exploitation, including the use of rock-lined roasting features for plant processing.

Although investigators have noted some indications of aboveground adobe structures at Middle Formative period sites in the Tularosa Basin, the few sites that have been excavated have tended to contain pit structures. In the beginning of the Middle Formative period, pit structures were often rectangular or subrectangular in plan view, suggesting a potentially greater investment in architecture and increased residential stability than would be indicated by circular or bean-shaped structures (Miller 2005). Differences in ceramic traditions at Middle Formative period settlements between the northern and southern Tularosa Basin suggest that the basin began to be divided into two culturally distinctive areas during that period or that ceramics were obtained through different exchange networks in the northern and southern portions of the basin (Vierra, Hanselka, and Windingstad 2010).

Recent excavations of Middle Formative period sites situated in the foothills of the Sacramento Mountains have recovered diverse plant remains, including agave, yucca, cheno-ams, pricklypear, strawberry cactus, cholla, grass, mesquite, and sotol (Miller et al. 2011), from burned-rock features. Foraging strategies appear to have involved the intensified exploitation of lower-ranked resources, such as cacti. Agave and cacti may have been used most intensively during the spring, when winter stores were exhausted, or throughout the year, during periods of resource shortfalls, such as times of drought. Indeed, the Middle Formative period is characterized by extreme shifts in precipitation that may have required adjustments to foraging and farming strategies for variation in precipitation. Subsistence activities are hypothesized to have focused on alluvial-fan and playa settings during wet years and the use of succulents in upland settings during dry years (Anderies et al. 2008; Dering 2005; Vierra 2012).

The Late Formative Period

The Late Formative period is marked by widespread changes throughout the greater Southwest, including sociopolitical and economic reorganization, population movement, and changes in architectural styles and iconography (Adler 1996; Clark 2001; Clark et al. 2006; Hill et al. 2004; Lyons 2003; Nelson 1999; Spielmann 1998). The timing and nature of these changes varied throughout the region and affected the various populations differently. In southern Arizona, platform mounds replaced ball courts as the primary form of communal architecture, and pit houses were replaced first by adobe-lined, semisubterranean pit rooms and later by adobe surface structures organized into room blocks. In the Tucson Basin, ceramic assemblages were dominated by Tanque Verde Red-on-brown pottery in the Late Formative A period, and polychrome styles were abundant during the Late Formative B period. Farther east, the Salado phenomenon emerged at that time, as indicated by the appearance of Salado iconography and architecture and the production of distinctively painted ceramic vessels (Crown 1994; Dean 2000). Many sites in the lower San Pedro Valley exhibit characteristics attributed to the Salado phenomenon, including adobe-and-cobble masonry architecture, single-story room blocks organized around a central plaza, and abundant Salado ceramics.

Evidence for the influx of northern groups from the Kayenta/Tusayan region during that time has also been recorded at a number of sites in the lower San Pedro Valley, at least 100 km north of Fort Huachuca (Clark et al. 2006; Di Peso 1958; Lindsay 1987; Lyons 2003; Woodson 1999). Sites located closer to Fort Huachuca, in the middle and upper San Pedro Valley, display characteristics that suggest that during the Late Formative B period, some groups in southeastern Arizona were influenced by the expansive Casas Grandes interaction sphere located to the south and east of Fort Huachuca (Altschul et al. forthcoming; Van West et al. 1997). Ceramic assemblages from that time are dominated by locally produced Babocomari Plain and Chihuahuan-inspired Babocomari Polychrome wares. Rectangular surface structures were constructed from puddled adobe or adobe-and-cobble masonry and were arranged around courtyards or plazas.

Late Formative period sites at Fort Huachuca include habitations, rock-feature sites, artifact scatters, and rock-art panels within rockshelters. The major Late Formative period site at Fort Huachuca is the Garden Canyon site, which had become a village by that time. Of the Late Formative period sites outside the fort, the most important is Babocomari Village (AZ EE:7:1 [ASM]), partly because it is the only Late Formative period site in the area that has been extensively excavated and fully reported (Di Peso 1951). Diagnostic artifacts at Late Formative period sites on Fort Huachuca include locally produced Babocomari Plain and Polychrome ceramics and imported (or emulated) ceramics, such as Gila Red, Belford Red, Tanque Verde Red-on-brown, Gila Polychrome, Santa Cruz Polychrome, and brown corrugated ware (Van West et al. 1997:179).

In the Tularosa Basin, reliance on maize agriculture appears to have reached its apogee during the Late Formative period. Late Formative period sites tended to be large and clustered in their distribution. High artifact densities suggest that sites were more intensively used than in earlier periods. Variability in architecture and site structure during the Late Formative period suggests differences in social organization and identity among sites. For instance, some villages consist of isolated or linear room blocks, and others consist of pueblos oriented around central plazas. In addition, some villages include structures built according to multiple architectural traditions. One interpretation of this variability is that Late Formative period sites in the Tularosa Basin represent multiethnic villages in which inhabitants from multiple Pueblo communities co-resided with each other (Kemrer 2006).

Late Formative period sites in the Tularosa Basin appear to represent a change in settlement distribution that coincides with an increased dependence on maize cultivation. These sites tend to be located in a variety of settings, including on alluvial fans, near playa depressions, and in upland locations (Carmichael 1986; Church et al. 2002; Kludt 2007; Leckman et al. 2009; Mauldin 1986). Several reservoirs dating to the Late Formative period have been discovered in the southern Tularosa Basin, indicating an increased investment in water control coincident with increased occupational intensity and maize cultivation (Leach et al. 1993; MacWilliams et al. 2009). A pattern of dual residences in upland and lowland locations may have also emerged during the Late Formative period (Kemrer 2008; Mauldin 1986). Vierra, Leckman, and Heckman (2010) have argued that the locations of Late Formative period fields varied with respect to annual rainfall: fields were located near playas and on alluvial fans during periods of high or average annual rainfall, on alluvial fans during periods of average rainfall, and at higher elevations in the Sacramento Mountains during periods of low annual rainfall.

Based on differences in ceramic types, Whalen (1978) has argued that occupation of the southern Tularosa Basin shifted from the eastern to the western side of the basin during the Late Formative period. Similarly, Lekson and Rorex (1994) observed that Late Formative period villages tended to be located in the foothills and canyons of the San Andres Mountains, on the western side of the basin. Overall, Late Formative period settlement represents a continuation of land-use strategies that appeared during the Middle Formative period (Vierra, Leckman, and Heckman 2010:13). By the end of the Late Formative period, upland groups began to move into the area of the San Andres Mountains, on the western side of the basin. Social conflict and warfare may explain these moves, because environmental pressures (e.g., drought) do not appear to have been a factor (Kemrer 2006; Wiseman 1997).

In addition, foraging groups moved into the region, followed by Athabaskan groups during the fifteenth century (Seymour 2002). The arrival of foraging groups may have resulted in a changing interaction sphere among local residents and more-recent immigrants. The Pueblo occupation of the Tularosa Basin appears to have ended by A.D. 1450, a change that may have been triggered by two extreme droughts occurring in A.D. 1405–1415 and ca. A.D. 1445–1450 (Grissino-Mayer et al. 1997). Drought may have resulted in increased competition over resources and resource depression, and possible evidence for warfare and violent conflict has been found at several sites (Hunter 1988; Miller and Graves 2009; Miller and Kenmotsu 2004; Seymour 2002).

Protohistoric and Historical Periods

Our understanding of the indigenous populations in southeastern Arizona and south-central New Mexico after the fifteenth century is limited and primarily comes from European historical documents. After ca. A.D. 1450, the settled villages that characterized the San Pedro Valley and the Tularosa Basin were depopulated. European explorers first ventured into southern Arizona in the sixteenth century, and early accounts indicated that the prehistoric populations of southern Arizona had been replaced by Upper Piman peoples. These groups established large settlements in the Santa Cruz and San Pedro River valleys and farmed the productive floodplains (Ravesloot and Whittlesey 1987; Seymour 1989). At some point during that time, bands of Chiricahua Apache expanded into the area, and Western Apache bands routinely traveled the region. These groups were highly effective raiders who viewed the agrarian populations of southern Arizona as an important subsistence resource (Basso 1983). By the end of the 1600s, Apache raids were common in the area, leading to frequent violent conflict and population movement. Diseases to which the indigenous populations had little resistance, introduced to the region by Europeans, also wreaked havoc on communities, causing devastating loss of life and undermining traditional subsistence practices and social systems (Reff 1991).

Major changes in settlement and social organization also appear to have occurred after A.D. 1450 in the Tularosa Basin, particularly with the arrival of mobile and aggressive Athabaskan groups in the region, but archaeological or documentary information is scant. Prior to the arrival of Spaniards in A.D. 1581, the Tularosa Basin was used most heavily by Mescalero Apache, whose territory centered on the mountain ranges surrounding the basin. Organized into relatively small bands based on kinship, the Mescalero Apache consisted of highly mobile groups whose subsistence depended on hunting, wild-plant collection, trade with agricultural Pueblo settlements, raiding of settlements for supplies and personnel, and some horticulture. Major food sources included buffalo meat obtained from the Plains as well as agave.

In the vicinity of the Tularosa Basin, Apache settlements depended on spring locations in the Organ, Sacramento, and San Andres Mountains surrounding the Tularosa Basin rather than within the basin itself. The basin was used mainly for hunting and gathering as well as for travel; moreover, the Apache threat appears to have kept other indigenous groups from using the basin (Basehart 1971; Faunce 2000; Opler and Opler 1950; Worcester 1941, 1979). The limited use of the basin, especially for residential purposes, left few traces in the archaeological record regarding the Apache presence.

The threat of Apache raiding activities was a severe impediment to use of the San Pedro River valley and the Tularosa Basin by Spaniards, Mexicans, and subsequent American settlers for the first several centuries of Euroamerican occupation of the region. While under control of the Spanish or Mexican governments, the Tularosa Basin was accessed mostly to obtain salt from Lake Lucero and other deposits, but otherwise, the basin was considered a desolate and largely uninhabitable area because of both the Apache threat and a lack of reliable water resources. Some livestock grazing in the southern portion of the Tularosa Basin and limited mining in nearby mountains occurred during the late eighteenth and early nineteenth centuries but still left few traces in the archaeological record because of fairly fleeting use of basin environments. Similarly, the San Pedro River valley was used to a limited extent during that time for agriculture, ranching, and mining, but settlements were often abandoned as a result of raiding activities.

With the acquisition of southern Arizona and southern New Mexico by the United States following the Mexican-American War and the treaty of Guadalupe Hidalgo, the San Pedro River valley and the Tularosa Basin began to be increasingly used for ranching, mining, and transportation as well as military activities. In both areas, ranches were established in many locations to exploit the extensive grasslands present at the time, and railroads were eventually established by the end of the nineteenth century.

Establishment of Fort Bliss and Fort Huachuca

In 1877, Camp Huachuca was established near the current location of the Old Post on Fort Huachuca, and it was upgraded to a fort in 1882 (Thiel 2005:6). The camp was established to protect settlers in the San Pedro and Santa Cruz River valleys from Apache raiding activities and to fulfill an agreement made in the Treaty of Guadalupe Hidalgo that the United States would prevent raiding parties from entering Mexico. The location of the camp, in a canyon along a well-watered creek, provided access to shelter and water and was considered a prime location for monitoring Apache movement (Rolak 1974). In the 1890s, black troops from the 24th and 25th Infantry and 9th Cavalry were temporarily stationed at Fort Huachuca for brief periods. The 10th Cavalry, another black unit, was moved to Fort Huachuca in 1913, and they were the only Army unit stationed at the fort until 1927. After 1927, the 10th Cavalry was joined by the 3rd Battalion, Headquarters, and Service Company of the 25th Infantry, another group of black troops. In 1931, Fort Huachuca ceased to operate as a cavalry post (Van West et al. 1997:300–301). During World War II, Fort Huachuca became a major training center for black troops (Smith 2001).

Fort Huachuca was deactivated in 1947. Control of the Fort was slated to be transferred to the State of Arizona for use by the Arizona National Guard and the Arizona Game and Fish Department in 1949, but the transfer was never finalized, although Arizona Game and Fish did run buffalo on the fort at that time. In 1951, Fort Huachuca was reactivated for the Korean War but was shut down again in 1953. In 1954, Fort Huachuca reopened under the control of the Army Signal Corps and was named the site of the Army Proving Ground (Trousil 2001:42; Van West et al. 1997:314–315). That same year, the Electronic Proving Ground was established on the fort. Thereafter, the Army assigned programs in military intelligence, research, and training to the post. In 1967, Fort Huachuca became the headquarters for the Army Strategic Communications Command, later renamed the Army Communications Command. The Army Communications Management Information Systems Activity was moved to Fort Huachuca in 1973. Eventually, the Army Communications Command and the Army Communications Management Information Systems Activity combined to form the Army Information Systems Command. In 1971, the Army Intelligence Center and School was established at Fort Huachuca (Trousil 2001:42–43). In 1974, the Old Post was listed in the NRHP as a district, and in 1976, the Fort Huachuca Historic District expanded and was named a National Historic Landmark. Today, Fort Huachuca is the headquarters for the Army Intelligence Center of Excellence and School and the Army Network Enterprise Technology Command/9th Signal Command, formerly the Army Signal Command (U.S. Fish and Wildlife Service 2007:4).

Fort Bliss Military Reservation had its beginnings in November 1848, when the U.S. War Department ordered the 3rd Infantry to take up quarters in El Paso, Texas. Originally called the “Post of El Paso,” the post was closed in September 1851, only to be reestablished in January 1854 at a new location 5 km (3 miles) east of the original post. A few months after being reestablished, the post was renamed “Fort Bliss,” after William Wallace Smith Bliss, a veteran of the Mexican War (Metz 1988:38–40). The fort was subsequently moved several other times before moving to its current location in 1893. In its early years, Fort Bliss was one of a series of forts located between Santa Fe, New Mexico, and San Antonio, Texas, that were intended to protect the expanding U.S. nation against hostile groups inhabiting the region. Although it was relatively small during the nineteenth century, troop strength at the fort was increased dramatically to 50,000 men in 1910, after the Mexican Revolution, and it was transformed from an infantry station to the largest cavalry post in the United States. The fort remained a post of the mounted U.S. Cavalry until 1943, when the U.S. Cavalry transitioned into using armored vehicles in place of livestock mounts. By that time, the fort was used primarily as an artillery post (Harris and Sadler 1993; Jamieson 1993; McMaster 1962).

Since World War II, the focus of most military activity at Fort Bliss has been on the training and testing of air-defense systems. The Anti-Aircraft Training Center was established by the Army at Fort Bliss in 1940, and by the end of World War II, Fort Bliss had established the nation's first guided-missile unit. Training and testing of air-defense systems continued at Fort Bliss during the Cold War and during the Gulf War of 1990–1991. As a result of the Base Realignment and Closure of 2005, the Air Defense Artillery Center and Air Defense Artillery School were moved from Fort Bliss to Fort Sill, Oklahoma. Fort Bliss was then transformed into a heavy-armor-training post. Today, the 1.1-million-acre reservation consists of guided-missile ranges and maneuver and training areas. The Main Cantonment is located adjacent to El Paso, Texas, but much of the reservation is in south-central New Mexico, as is the project site.

Project Sites

In consultation with the cultural resource managers for the two installations participating in this project, Fort Bliss and Fort Huachuca, two sites were selected for investigation, one at each installation. The site located on Fort Bliss Military Reservation, FB 9583, was one of three sites (FB 9583, FB 9617, and FB 18878) investigated recently as part of a mitigation project on Fort Bliss. The goal of the project was to minimize adverse effects on the three sites by recovering data that addressed the research questions presented in the plan of work (Vierra et al. 2011). FB 9583, which had been discovered during a previous survey (Hanselka et al. 2010), was selected for inclusion in the current project because it had a suitable number and diversity of ceramic and lithic artifacts. The mitigation project that included investigation of FB 9583 dovetailed nicely with the current project, allowing the in-field and laboratory analyses to be performed in tandem with the mitigation efforts without affecting the schedule or cost of the mitigation project.

The other site investigated for the current project was the Soldier Creek site on Fort Huachuca. The Soldier Creek site had been recorded during two previous surveys (Cook 2004; Vanderpot 1994:86–92) and was selected by the cultural resource managers as a site that would have been especially important to interpret accurately for management concerns and could produce a suitably large and diverse collection of ceramic and lithic artifacts for analysis.

FB 9853

FB 9853 is a large artifact scatter with features interpreted as a prehistoric campsite that is situated on a middle to late Pleistocene alluvial terrace along the western bank of El Paso Draw, in a location where a canyon floor expands to form a broad valley (Hanselka et al. 2010:127–130) (Figure 2). The site is located in the foothills of the Sacramento Mountains, on the eastern side of the southern Tularosa Basin, along the edge of Otero Mesa, a vast grassland that has been described as rivaling the Serengeti in the lushness of its grassland ecosystem (New Mexico Wilderness Alliance 2013). The local environment surrounding the site is characterized by rocky limestone hills divided by small drainages and arroyos. Vegetation is dominated by grasses, especially black grama and blue grama. Creosotebush, mesquite, soaptree yucca, Torrey's yucca, and cholla are also relatively common. Dense stands of desert willow, javelina bush, littleleaf sumac, and mesquite are found in riparian areas associated with El Paso Draw. Sediments at the site are characterized by calcareous alluvium and colluvium/slopewash. Pockets of alluvium, sometimes as much as 5 m deep, are exposed at the site along the steep banks of El Paso Draw. Erosion resulting from several small rills that drain the surrounding hillsides has affected deposits in the northern and western parts of the site. Other disturbances to the site include a historical-period two-track road that runs along the northern end of the site and along the western portion of the site.

Figure 2. The physical environment of FB 9853, Fort Bliss Military Reservation, New Mexico, view to the northwest.

FB 9853 was originally documented by Beckes et al. (1977), who described the site as a large occupational area with more than 50 fire-cracked-rock (FCR) concentrations and middens. More recently, Hanselka et al. (2010) described FB 9853 as a multicomponent campsite and possibly a seasonal residential site with approximately 100 FCR middens and concentrations and a single stain. Features were clustered in the southern half of the site. At least 20 of the features were identified as containing potentially datable fill. Hanselka et al. (2010) identified more than 500 ceramic and lithic artifacts at the site; diagnostic artifacts indicated that the site was used during the Late Archaic and Formative periods. It was interpreted as being in an ideal location for runoff farming and was recommended eligible for listing in the NRHP on the basis of its setting, the large numbers of datable artifacts, and the likely presence of intact buried features and datable deposits.

The mitigation effort conducted in 2012 revealed the presence of 34 FCR middens, 63 FCR concentrations, and 1 stain interpreted as an occupation surface (Vierra and Ward 2012). As part of the mitigation project, 10 of the originally documented FCR middens and concentrations were partly excavated. In addition, the occupation surface was exposed and excavated, and 8 pit subfeatures were partly or fully excavated. Radiocarbon dates obtained on materials excavated from the site indicated site use during the Formative period, between 1510–1350 and 650–540 cal yr B.P.

To relocate the site, SRI staff used a mapping-grade Global Positioning System (GPS) receiver loaded with geospatial data representing the locations of the previously recorded site boundaries, datums, and features. Once these attributes were relocated, the GPS unit was used to establish Universal Transverse Mercator North American Datum 1983 coordinates for new site-mapping datums and backsight locations. A total station was used for site mapping. All point-provenienced surface-collection locations corresponding to the artifacts analyzed for the current project were mapped, using the total station and established site datums (as were all backhoe trenches, hand-stripping units, excavation units, surface-collection units, features, subfeatures, mapping nails, and important modern or natural features).

Flaked stone, ground stone, and identifiable decorated ceramics at FB 9583 were point-plotted and collected from across the site. Data from the previous SRI survey of the site (Hanselka et al. 2010) were used to locate the recording units in which these artifacts had previously been identified on the site surface. The recording units were then field inspected, and items were collected. Collected artifacts were bagged by unit and separated into general material categories, such as flaked stone and ceramics; each artifact was assigned a single provenience designation (PD) number so that the identifications of individual artifacts resulting from separate in-field and laboratory analyses could be directly compared.

Lithic artifacts collected from the site, a sample of which was analyzed as part of this project, included cores, debitage, retouched tools, and ground stone tools. Most lithic artifacts were made on chert, and around one-fifth of the artifacts were made on limestone. In rare cases, lithic artifacts were made on chalcedony, obsidian, quartzite, rhyolite, or sandstone. Ceramic artifacts from the site indicated that the site was occupied during the Middle Formative (A.D. 1000–1275/1300) and/or Late Formative (A.D. 1275/1300–1450) period. El Paso Brown rim sherds and sherds associated with El Paso Bichrome and nonlocal wares placed the site within the Middle Formative period; however, El Paso Polychrome sherds, which were most common during the Late Formative period, represented more than half (57 percent) of the diagnostic ceramics from the site. Evidence of food resources from macrobotanical and pollen analyses included pricklypear (*Opuntia* spp.) and hedgehog cactus (*Echinocereus* spp.) seeds, maize (*Zea mays*) cupules and pollen from maize, squash (Cucurbitaceae), cattail (*Typha*), cholla (*Opuntia* spp.), pricklypear, cactus family (Cactaceae), lily family (Liliaceae), grasses, and cheno-ams (*Chenopodium*/*Amaranthus*). Faunal remains from the site included those of cottontails (*Sylvilagus* spp.), jackrabbits (*Lepus* spp.), artiodactyls, and carnivores as well as a fragment of marine shell.

The Soldier Creek Site

The Soldier Creek site was a large, intensively utilized Formative period habitation site located near the end of a broad ridge that overlooks a wide, grass-covered floodplain of Soldier Creek in southeastern Arizona (Figure 3). The site was originally recorded by Vanderpot (1994) during survey and was rerecorded during a subsequent survey by Cook (2004). Sediments at the site consist of mixed gravel and loamy alluvium. Vegetation is dominated by mesquite; desert broom, burroweed, and cholla are also present. The site consists of an extensive ceramic and lithic scatter, five trash mounds, and two rock-ring features. An isolated historical-period trash scatter and modern artifacts associated with military activity are also present at the site.

Lithics recorded at the site during previous investigations consisted mostly of debitage and cores made on a variety of raw materials, including chert, jasper, hornfels, quartzite, quartz, chalcedony, rhyolite, and andesite. One piece of obsidian has also been observed. Flaked stone tools previously identified at the site included modified flakes, a scraper, core/chopper tools, tool blanks, and two projectile points. Ground stone artifacts previously identified at the site—all of which were made on granite, quartzite, or an unidentified material—consisted of at least six fragments from basin metates, a trough-metate fragment, and four manos. Ceramics consisted of a variety of painted and plain wares, including wares associated with the Tucson Basin, Dragoon, and Babocomari ceramic traditions (Heckman et al. 2000). Although pit-house features are suspected to be present, based on the presence of trash mounds, they have not been observed on the site surface.

When first recorded by Vanderpot (1994), thousands of artifacts were noted at the site, including hundreds on the surfaces of individual trash mounds. When rerecorded by Cook (2004), however, the surface characteristics of the site had changed; only 110 flaked stone artifacts, 102 ceramic artifacts, and 7 ground stone artifacts were noted. Erosion was inferred to have caused changes to the site surface between recording episodes. In addition to erosion, the site has been impacted by vehicle traffic and modern military activity; at least one feature (Feature 2) has been directly impacted by modern activities. Multiple collections have also been made from the site, including a substantial number of decorated sherds. Visits to the site by the Fort Huachuca cultural resource manager in July 2011, at a time when visibility was excellent,

Figure 3. Field crew recording artifacts at the Soldier Creek site (AZ EE:7:164 [ASM]), Fort Huachuca, Arizona, view to the northwest.

revealed that the number of artifacts visible on the surface of the site had increased since Cook's recording of the site in 2002, and diverse artifacts were visible on the surfaces of trash mounds at the site. The site was thus considered to be a viable candidate for the current study.

Permission to conduct investigations and collect artifacts from the site in support of the current project involved extensive consultation with the Advisory Council on Historic Preservation (ACHP), the Arizona SHPO, and other relevant stakeholders, including Native American tribes. To initiate consultation, Fort Huachuca discussed with the Arizona SHPO the potential for artifact collection at the Soldier Creek site, and the Arizona SHPO identified artifact collection as an adverse effect. This finding suggested the need for consultation with Native American tribes with an interest in historic properties on Fort Huachuca. Of the tribes consulted, the Tohono O'odham Nation (TON) registered concerns about collecting artifacts from a site that was not being disturbed. As an adverse effect, artifact collection as originally proposed for the project would have required the drafting of a Memorandum of Agreement (MOA) and associated input from the ACHP. Consultation regarding these concerns led to agreement on a revised strategy of returning the artifacts to the site after laboratory analysis. This strategy eliminated the adverse effect for the Arizona SHPO and the TON and, by extension, eliminated the need for the MOA or for ACHP input. This led to a work plan detailing how the project was to be conducted at the Soldier Creek site (see Appendix A). The scope of work indicated that each of the trash mounds would be sampled, because the number and diversity of artifacts in each trash mound could potentially be limited. However, when field investigations were conducted for the current project, an adequate number and diversity of artifacts were observed as present on the surfaces of the trash mounds, making sampling of each of the four trash mounds unnecessary.

The ACHP and the Arizona SHPO agreed that if artifacts collected during the course of the project were returned to the field by the end of the project, rather than curated, there would be no adverse effect on the Soldier Creek site. Thus, the collected artifacts were analyzed and photographed in the field and subsequently analyzed in the laboratory before being returned to their original collection units. To ensure that collected artifacts were returned to the site in close proximity to their original positions in the field, artifacts were point-provenienced, and photographs, each including a metric scale and a north arrow, were taken of the surface contexts and the positions of the investigated artifacts prior to their collection.

Methods

Field recording of artifacts was performed using SRI's system for recording provenience information for artifacts, features, and sample units. Following widely accepted conventions for identifying the discovery locations of artifacts, samples, and activity areas, this system provides a separate list of PD numbers for each archaeological site and assigns a unique number to every space that contains features, artifacts, or other archaeological units. For this project, all of the PD numbers of the collected artifacts and their associated in-field and laboratory observations were entered into SRI's relational database (SRID). This facilitated comparison between the in-field and laboratory analyses of both the physical artifacts and the digital photographs of the artifacts.

Digital photographs were taken of the obverse and reverse surfaces of each artifact (e.g., the interior and exterior surfaces of a ceramic artifact) (Figures 4 and 5); a scale was included in each photograph. All photographs were documented in a photograph log, such that each individual artifact could be clearly identified and related to its analysis data.

In total, 20 ceramic artifacts and 105 lithic artifacts from FB 9583 and 345 ceramic artifacts and 150 lithic artifacts from the Soldier Creek site were analyzed. The artifacts analyzed from the Soldier Creek site included 2 projectile points and 11 ceramic sherds that had been collected from the site prior to the project and curated by either the Arizona State Museum (ASM) or Fort Huachuca. All ceramic materials collected during the project were classified into traditional ware and type categories, to the finest level possible. Lithic analysis focused on the identification of artifact types, raw-material types, and technological attributes. In some cases, specific quantitative methods were used to calculate measures, such as estimated vessel counts or the precision of temporal affiliations. Because these methods are specific to either the ceramic or the lithic analysis, they are presented in the chapters that assess the results of the ceramic and lithic analyses (Chapters 3 and 4, respectively).

Each artifact to be collected was identified in the field by a field technician, who identified the artifact according to type and relevant technological attributes, depending on material class. The same artifacts were analyzed a second time by another field technician, using the provenience information and artifact identifiers established during the first recording. In the rare instances in which questions arose regarding which artifacts pertained to particular PDs, digital photographs and photograph logs were reviewed to ensure that the correct artifacts were analyzed. After being analyzed a second time in the field, the artifacts were collected; standardized procedures for the collection and storage of artifacts (in accordance with 36 CFR 79 and established curatorial standards and guidelines) were followed.

Laboratory Artifact Analysis

Laboratory processing followed established SRI guidelines that meet all the requirements of the ASM and the Directorate of Public Works, Environmental Division, Conservation Branch, of Fort Bliss (Fort Bliss 2012). Archival-quality storage materials were used. Computerized inventories of artifacts, proveniences, and catalog lists were maintained by the laboratory. Computer files, databases, and inventories were regularly backed up, and copies of all records were kept in fireproof facilities. The laboratory analyses and recording of collected artifacts mirrored the in-field recording.

The physical artifacts, along with digital photographs of the artifacts, were analyzed in the laboratory by separate analysts. A ceramic analyst analyzed only the digital photographs of the ceramic artifacts collected

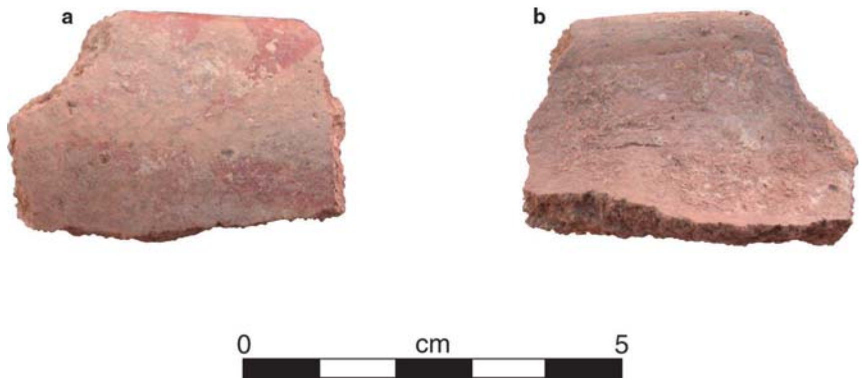

Figure 4. Digital field photographs of the (*a*) interior and (*b*) exterior surfaces of a Dragoon Red-on-brown ceramic rim sherd documented during fieldwork at the Soldier Creek site (AZ EE:7:164 [ASM]), Fort Huachuca, Arizona.

Figure 5. Digital field photographs of the (*a*) obverse and (*b*) reverse surfaces of a chert biface documented during fieldwork at FB 9853, Fort Bliss Military Reservation, New Mexico.

from each site; a lithic analyst analyzed only the digital photographs of the lithic artifacts collected from each site. In both cases, the digital photographs were displayed on computer monitors during analysis. The physical ceramic and lithic artifacts were analyzed separately, in a laboratory setting, by a different ceramic analyst and a different lithic analyst, respectively. To ensure independence of analyses, the analysts who examined only the digital photographs were not provided the opportunity to examine the physical artifacts or to revise identifications after the digital-photograph analysis had been completed.

Methods for Assessing Artifact-Identification Results

In assessing the artifact-identification results, it is important to recognize that different ceramic specialists performed the hands-on and digital-photograph laboratory analyses, and different field technicians performed each of the in-field analyses, as discussed above. Field technicians and ceramic specialists made their observations independently of other analyses and were not permitted to view the results of other analyses of the same collected artifacts.

Another important factor to consider is that the artifact types and attributes identified during the project all conformed to the same typological systems. Descriptions of specific artifact types and methods of identification are described in technical field manuals that were provided to the analysts prior to and during fieldwork and analysis. Thus, discrepancies among analyses in artifact identification did not result from the application of different typological systems but reflect different choices made by the analysts in placing an artifact within a standardized, defined, and mutually recognized typological category.

Another factor to consider is that all identifications were entered into a sophisticated relational database system (SRID) for which data entry is tightly controlled. Data entry in SRID is funneled from less-specific attributes and observations to more-specific observations. For example, if the user selects "flaked stone" from the initial drop-down list, then the remaining observations, attributes, and choices on the drop-down lists are limited to only those observations and attributes relevant to a flaked stone artifact. The entered data are also relationally tied to other attribute data, such that associated attributes (e.g., the associated date range for a specific ceramic type) can be filled in automatically, based on related attribute information; logically inconsistent data entries are not permitted, such as identifying the artifact class of an artifact typed as a mano as "flaked stone artifact" rather than "ground/battered stone artifact." Data entry was quality controlled to ensure that mistakes or transcription errors, such as entering the wrong attributes for a specific PD number, were prevented. All of these factors helped to ensure that discrepancies in the artifact-attribute data between analyses actually reflect discrepancies in artifact classification rather than data-entry mistakes or differences in typology, attribute definitions, or labels.

A final issue of importance in assessing the level of agreement among the artifact analyses conducted for the project is that the identifications result in the correspondence of categorical data to specific artifact types or attributes. None of the data collected as part of the artifact analysis are continuous measurements, such as linear measurements of an artifact's dimensions. Thus, the approaches applied to evaluating the results were based on methods for assessing interobserver variation in categorical, rather than continuous, observations. Statistics developed for assessing interobserver variation in continuous or ratio-scale measurements, though more commonly applied in assessing interobserver variation, could not be usefully applied.

Analytical Approach

In general, discrepancies among observers making scientific observations of the same objects or phenomena are to be expected. Discrepancies can result from instrumentation error, variation in the application of measurement or observation techniques, mistakes, observer bias, or other factors. Observer bias can result from different or imprecise definitions of observational categories, differences in interpretation or perception, or variation among observers in experience or training. Differences between sets of observations can also occur when observations are made at different times by the same investigator (as a result of a subtle shift over time in the definitions or methods applied in making observations) or as a result of changes in instrumentation or environmental conditions (Adams and Adams 1991; Beck and Jones 1989; Brennan and Silman 1992; Caro et al. 1979; Dibble and Bernard 1980; Fish 1978; Lyman and VanPool 2009; Majewski and O'Brien 1987; Neuman et al. 1999; Whittaker et al. 1998).

Interobserver variation is traditionally divided into three categories: random variation, systematic variation, and illegitimate variation (Beers 1957; Daniels 1972). Random variation is generally described as occurring as a result of unpredictable variation in the conditions of an experiment, such as changes in environmental conditions or measurement instruments. For ratio-scale variables, random errors should have no effect on the central tendency of a measurement but will effect the variance of a measurement (with more random error resulting in a larger variance). In the context of the current project, random error could have resulted from such factors as variation in the environmental conditions in which observations were made, variation in the quality of digital photographs, or random variation in the interpretation or application of typological definitions.

Systematic variation occurs when one set of observations is consistently offset from another set of observations of the same objects or phenomena. Systematic variation occurs as a result of instrument errors or when one observer consistently applies a typological or measurement scheme in a manner that is different from how it is applied by another observer. For instance, if a measurement instrument is set incorrectly, length observations made with that instrument could be consistently too long or too short. Unlike random errors, systematic errors can affect the central tendency of a measurement. For the analysis of artifact types, systematic error could occur when one observer consistently identifies one artifact type incorrectly as another type or types, such as the case in which one observer consistently identifies artifacts made on chert as having been made on limestone.

Illegitimate variation occurs as a result of unintentional mistakes, such as making observations of the wrong artifact or transcribing data incorrectly into an analysis table. Investigators typically attempt to weed out or remove from consideration illegitimate errors in the assessment of interobserver variation. Such variation tends to produce outliers or anomalies in the data that correspond more to outright mistakes than to the level of agreement between observers. The potential for illegitimate variation was minimized in the current project by applying a controlled PD system and a series of controls for preventing artifacts from being mixed up or data from being incorrectly entered into the database system.

In scientific investigations, differences among observers become important when they are substantial enough to result in a significantly different interpretation of the materials or phenomena being investigated. In terms of artifact analysis, differences between observations are most relevant when they affect the interpretation of a site or behavioral pattern. For instance, discrepancies between observations that result in differences in the inferred cultural or temporal affiliation of a component or the kinds and relative intensities of activities performed at a site can adversely affect how a site is interpreted. For CRM, differences in site interpretation can influence management decisions and eligibility recommendations. An important site may be considered relatively unimportant, for instance, if incorrectly interpreted, or it could potentially be placed in a management category that is inappropriate to effective management of the site. It is thus of utmost importance that managers are able to assess the accuracy and reliability of site interpretations in order to make informed management decisions.

Although potential problems with variation between observers in archaeological-artifact analysis and description have been raised by a number of investigators, the development or implementation of quantitative methods for evaluating interobserver variation in archaeology has been limited. Assessment of interobserver variation has been more common in other disciplines, including behavioral ecology (Caro et al. 1979), geology (Friedman 1958; Mizutani 1963), physical anthropology (Gavan 1950; Jamison and Zegura 1974), and medicine (Brennan and Silman 1992; Shoukri 2010). Of the few archaeological studies that have used quantitative methods for comparing artifact-analysis results, some of the more-intensive studies have focused on evaluating interobserver variation in ratio-scale or continuous variables, such as projectile point length or width, rather than on typological categories (e.g., Dibble and Bernard 1980; Lyman and VanPool 2009). Exceptions include studies reported by Fish (1978), Swarthout and Dulaney (1982), Boyd (1987), Beck and Jones (1989), Whittaker et al. (1998), and Gnaden and Holdaway (2000).

Uniquely, Gnaden and Holdaway (2000) assessed the reliability of in-field artifact analysis performed by field technicians and in-field analysis performed on the same artifacts by a trained specialist. Beck and Jones (1989) investigated the potential for bias between different sets of field observations recorded during survey, but they did so by comparing results derived from two analysts who analyzed different sets of artifacts from different contexts (off-site and on-site contexts). Other studies have been designed to evaluate the efficacy of typologies or classification systems by assessing the level of agreement in artifact classification among trained specialists in a laboratory setting.

Measurement of Interobserver Variation

For ratio-scale data, interobserver variation is usually assessed (in archaeology) using statistics aimed at testing whether the distribution of measurements made of the same artifacts by different observers or using different observational methods are significantly different. Often, such studies test for statistically significant differences in the mean and variance of a measurement, using statistical approaches such as *t*-tests or analyses of variance (Lyman and VanPool 2009).

For typological data, quantitative evaluation of interobserver variation is often accomplished by calculating the percent or proportion of observational cases in which two sets of observations agreed or disagreed. For instance, Fish (1978) briefly presented a study aimed at assessing the level of interobserver agreement among four ceramic analysts in identifying the ceramic type of 90 individually numbered Kayenta Tusayan gray ware and white ware ceramic artifacts. Kayenta Tusayan gray ware and white ware ceramic artifacts were chosen for the study because they were considered to be some of the better-studied and clearly defined prehistoric ceramic types in the Southwest. The four observers who participated in the study were chosen because they all had experience in identifying the ceramic types represented in the collection and were trained by the same specialist in the same institutional setting.

This study showed that the level of agreement between any two sets of participating observers ranged from 69 to 78 percent. No individual ceramic type was consistently identified differently by observers. However, when disagreements occurred between analysts, one analyst consistently identified ceramic artifacts as corresponding to a type that dated *earlier* than the type identified by other analysts. Another analyst consistently identified ceramic artifacts as corresponding to a type that dated *later* than the type identified by other analysts. The study suggested that substantial random variation and systematic variation can occur between analyses of the same artifact types, even when analyses are performed in a laboratory setting, by similarly trained specialists, and on artifacts corresponding to clearly defined and widely recognized artifact types.

A somewhat more sophisticated means of calculating the level of agreement between typological analyses can be achieved through the use of contingency tables to calculate an agreement statistic developed for categorical data, referred to as Cohen's kappa (Brennan and Silman 1992; Cohen 1960; Gnaden and Holdaway 2000; Shoukri 2010; Sim and Wright 2005). An advantage of Cohen's kappa is that it not only takes into account the observed agreement, but it also takes into account the agreement that might be expected as the result of chance. Another metric, referred to as McNemar's test for bias, can be calculated to test for systematic variation in categorical observations.

When one set of observations is treated as a "gold standard," another means of assessing agreement between observers that can be usefully applied is to calculate false-positive and false-negative rates in a manner analogous to how such rates are calculated in testing the performance of a statistical model. Such rates allow the assessment of which observers and observational methods tend to identify an artifact type as present when the artifact type is absent (false positive) as well as which observers and observational methods tend to identify an artifact type as absent when it is actually present (false negative).

Below, we discuss five metrics used to assess interobservation in this study: an agreement index, a false-positive rate, a false-negative rate, Cohen's kappa, and McNemar's test for bias. This is followed by a brief discussion of qualitative and interpretive assessment methods applied during the project.

The Agreement Index

The agreement index provides the proportions of artifact identifications that are in agreement with each other for a given pair of analyses (e.g., in-field analysis performed by a field technician and digital-photograph analysis performed by a trained specialist). This is one of the more common metrics used to assess the level of agreement between observations for categorical variables (see Caro et al. 1979). Because no assumptions are made regarding the accuracy of artifact identifications in calculating the agreement index, it provides an unbiased estimate of which pairs of analyses were in greatest or least agreement with each other for a given artifact attribute (e.g., ceramic ware, material type, or artifact type). The index varies from 0 to 1; 0 indicates no agreement, and 1 indicates perfect agreement. The agreement index was calculated per project for each pair of analyses in order to compare the level of agreement among the four sets of analyses: the in-field analyses performed by each of two field technicians, the digital-photograph analysis performed by a trained specialist in a laboratory setting, and the hands-on analysis performed by a separate trained specialist in a laboratory setting. The result was a series of six interobserver comparisons per attribute and project site. For ceramic artifacts, the agreement index was calculated for the following attributes: vessel element, ceramic ware, and ceramic type (see Chapter 3). For lithic artifacts, the agreement index was calculated for the attributes of production method (e.g., flaked stone, ground/battered stone, or expedient use [e.g., hammerstones and hand stones]), material type, and technological type (e.g., biface, flake, or core) (see Chapter 4). An overall level of agreement was also derived by calculating, between sets of observations, the proportion of all observations that were in agreement with each other at each of the two project sites.

Misidentification Rates: False Positives and False Negatives

Another two metrics calculated to assess interobserver variation in artifact identification for this project were the false-positive rate and the false-negative rate. These rates were calculated under the assumption that the most accurate classification was made during the hands-on laboratory analysis, by a trained specialist. In other words, for these calculations, the hands-on laboratory analysis was treated as the "gold standard" of artifact identification (see Carlin 1993; Gnaden and Holdaway 2000:740), and the results of the other analyses were measured against this standard. An artifact that was correctly identified by an analysis, based on the results of the hands-on laboratory analysis, was considered to be a true-positive case for that analysis. For a given artifact type, a false negative represents a case in which an artifact type (as identified during the hands-on laboratory analysis) was not identified correctly. A false positive represents a case in which the hands-on analysis did not identify the particular type for a specific artifact but another analysis did. In other words, a false negative indicates a classification that was missed (e.g., the hands-on laboratory analysis identified an artifact as a biface, but another analysis failed to identify the artifact as a biface). A false positive indicates a classification that was falsely made (e.g., an artifact not identified as a biface by the hands-on laboratory analysis was identified as a biface by another analysis).

The false-negative rate is analogous to a Type I error in statistics, or the assertion that something is absent when it is actually present. For the purpose of this project, the rate was calculated, according to type, as the number of false-negative cases divided by the number of hands-on-analysis cases. In other words, if the hands-on analysis identified four artifacts as bifaces, but only one of those artifacts was identified by the digital-photograph analysis as a biface, then the false-negative rate for bifaces for the digital-photograph analysis would be 0.75 (the three bifaces misidentified as another artifact type by the digital-photograph analysis divided by the four bifaces identified by the hands-on laboratory analysis). A false-negative rate of 0 indicates that an artifact type identified by the hands-on laboratory analysis was *never* misclassified (classified as a different type) by another analysis. A false-negative rate of 1 indicates that an artifact type identified by the hands-on laboratory analysis was *always* misclassified (classified as a different type) by another analysis. The rate cannot exceed a value of 1, because there can never be more false-negative cases than there were total cases identified by the hands-on analysis.

The false-positive rate is analogous to a Type II error in statistics, or the assertion that something was present when it was actually absent. The false-positive rate indicates the rate at which an artifact was falsely identified as being of a particular type when it was actually a different type, based on the results of the hands-on laboratory analysis. For the purpose of this project, the rate was calculated, according to type and analysis, as the number of false-positive cases divided by the number of hands-on-analysis cases. In other words, if the hands-on analysis identified four artifacts as bifaces, as in the above example, but an additional seven artifacts were identified by the digital-photograph analysis as bifaces (and were identified as different artifact types by the hands-on laboratory analysis), the false-positive rate would be 1.75 (the seven artifacts misidentified as bifaces by the digital-photograph analysis divided by the four bifaces identified by the hands-on laboratory analysis). A false-positive rate of 0 indicates that, for a given type, an analysis did not identify any additional cases of that artifact type that were not identified by the hands-on analysis. A false-positive rate of 1 indicates that, for every artifact of a given type identified by the hands-on analysis, another analysis misidentified an equal number of additional artifacts as that artifact type. A false-positive rate exceeding 1 indicates that more artifacts were erroneously identified as a particular type than were actually present in the collection.

In this report, in order to compare the false-positive and false-negative rates among analyses, we consider false-negative rates and false-positive rates of 0.25 or below to be low. Rates between 0.26 and 0.5 are considered moderate, and rates above 0.5 are considered high.

Cohen's Kappa and McNemar's Test for Bias

For some of the more common artifact types identified at the two project sites, two additional metrics were calculated: Cohen's kappa and McNemar's test for bias (Bishop et al. 1975; Brennan and Silman 1992; Gnaden and Holdaway 2000). The statistics were not calculated for types that were not identified by the hands-on analysis, types that were identified only for a handful of artifacts, or types for which most artifact identifications were made by a single observer. For these situations, either the tests are unnecessary as raw counts, and other statistics clearly indicate poor agreement, or sample sizes are especially small.

Cohen's kappa is a statistical measure of agreement for categorical data calculated using a contingency table in a manner similar to calculation of the chi-square statistic (Table 1). The contingency table is constructed to indicate the number of cases in which two observers agree in identifying whether an artifact is or is not a particular type (e.g., both agree that an artifact is a core or that an artifact is not a core) as well as the number of cases in which only one of the observers identifies an artifact as a particular type (e.g., one observer identifies an artifact as a core, and another observer identifies the artifact as a type other than a core). Like the false-positive and false-negative rates discussed above, Cohen's kappa and McNemar's test for bias were calculated under the assumption that the hands-on laboratory analysis was most likely to have resulted in accurate classifications. Thus, comparisons were made between the hands-on laboratory analysis and the other analyses in order to assess the level of agreement of the latter group of analyses with the "gold standard" for the project.

In Table 1, variable *a* represents the number of cases in which *both* observers identify an artifact as the type for which the comparison is made (e.g., core flake); variable *d* represents the number of cases in which *both* of the observers agree that an artifact is *not* the type for which the comparison is made. Variable *c* represents the number of cases in which *only* Observer 2 identifies an artifact as the type under consideration, and variable *b* represents the number of cases for which *only* Observer 1 identifies an artifact as the type under consideration.

For instance, consider the situation in which the level of agreement is assessed between observers for artifacts typed as core flakes. If there are 100 artifacts, and both Observers 1 and 2 identify 20 of the artifacts as core flakes and 40 as *not* core flakes, then *a* would equal 20, and *d* would equal 40 (Table 2). If, of the remaining 40 artifacts, Observer 1 identifies 10 artifacts as core flakes (and 30 as *not* core flakes), and Observer 2 identifies 30 artifacts as core flakes (and 10 as *not* core flakes), then *b* would equal 10,

Table 1. Matrix Used for Calculating Cohen's Kappa Statistic

	Observer 2 Yes	Observer 2 No	Total
Observer 1 yes	*a*	*b*	g_1
Observer 1 no	*c*	*d*	g_2
Total	f_1	f_2	*n*

Note: After Gnaden and Holdaway (2000:Table 1) and Sim and Wright (2005:Table 1).

Table 2. Example Matrix for Calculating Cohen's Kappa Statistic to Assess Agreement in Core-Flake Identification between Observers

	Observer 2 Yes	Observer 2 No	Total
Observer 1 yes	20	10	30
Observer 1 no	30	40	70
Total	50	50	100

and *c* would equal 30. The total number of artifacts, 100, would be *n*. The column totals, f_1 and f_2, would each be 50; the row totals, g_1 and g_2, would be 30 and 70, respectively (see Table 1).

Using the above variables, the observed (A_o) and expected (A_e) values for the number of cases in agreement with each other can be calculated. The number of observed agreement cases is simply *a* plus *d*. In the example provided above, the observed agreement equals 60 (20 + 40). The expected number of agreement cases can be calculated using the following formula:

$$A_e = ([f_1 \times g_1]/n) + ([f_2 \times g_2]/n) \quad (1.1)$$

In the above example, the expected agreement would equal 50 ([(30 × 50)/100] + [(70 × 50)/100]), or 10 less than the observed agreement.

Using the observed and expected agreement and the total number of artifacts, Cohen's kappa (κ) can be calculated as follows:

$$\kappa = (A_o - A_e) / (n - A_e) \quad (1.2)$$

Cohen's kappa varies from –1 to 1. Values above 0.8 indicate very good agreement, and values below 0.2 indicate poor agreement (see below). Using the above example, Cohen's kappa would equal 0.2 ([60–50]/[100–50]), a low value, indicating poor agreement. To put this in perspective, if Observer 1 conducted the hands-on laboratory analysis, then the false-negative rate would be 0.33 (10/30), because out of a total of 30 core flakes identified by Observer 1, Observer 2 failed to identify 10 of these artifacts correctly. The false-positive rate would be 1 (30/30), because Observer 2 identified an additional 30 artifacts as core flakes that were not actually core flakes. Thus, although one might consider the false-negative rate to have been moderate in the example, the false-positive rate was quite high, resulting in a low level of agreement overall, as indicated by Cohen's kappa.

Brennan and Silman (1992:Table 2) have provided a table indicating the suggested interpretation for different values of Cohen's kappa (Table 3). In addition to applying Brennan and Silman's (1992) levels for interpreting the strength of agreement between observers, a threshold value can also be assigned for assessing whether the level of agreement is acceptable or unacceptable. Although many applications of the metric have used values of 0.4 or above to indicate an acceptable level of agreement, Gnaden and Holdaway (2000) interpreted Cohen's kappa values equal to or above 0.5 as representing an acceptable

Table 3. Interpretation of Agreement Strength for Different Values of Cohen's Kappa

κ Statistic	Strength of Agreement
<0.2	poor
0.21–0.4	fair
0.41–0.6	moderate
0.61–0.8	good
0.81–1	very good

Note: After Brennan and Silman (1992:Table 2).

level of agreement; they interpreted values less than 0.5 as representing an unacceptable level of agreement. To be consistent with Gnaden and Holdaway's (2000) analysis, which also evaluated the reliability of in-field artifact analysis, a cutoff value of 0.5 for Cohen's kappa to indicate an acceptable level of agreement has been adopted for this report.

A companion test that can be calculated using the same contingency table used to calculate Cohen's kappa is McNemar's test for bias. This test can be used to identify whether systematic bias has occurred in making a typological classification, based on the chi-square distribution. In other words, if one observer frequently identifies artifacts as a particular type, and another observer frequently identifies the same artifacts as a different type, the McNemar's test for bias can be used to test whether the difference in identification represents systematic bias or random error. To perform McNemar's test for bias, a chi-square value is calculated as follows:

$$([b - c] - 1)^2 / (b + c) \tag{1.3}$$

In this report, a chi-square value of 3.84 or above is used to identify the presence of systematic bias at the 95 percent confidence level. Using the above example, the chi-square value would be calculated as 11.025 ($[(10 - 30) - 1]^2 / [10 + 30]$), which strongly indicates that systematic bias occurred between these hypothetical observers in their identifications of whether or not the artifacts were core flakes.

Qualitative and Distributional Methods

A final means of evaluating artifact-analysis results was to evaluate how different the interpretations of the project sites would be, had one analysis been used instead of another. For these comparisons, it is not important that the identification of an individual artifact was the same or different between analyses. What is important is whether the overall distribution of artifact types resulting from the analyses was consistent or inconsistent among observers. In these assessments, particular attention was paid to those factors that could have the greatest potential to influence the interpretation of a site, such as whether the analysis results could be used to infer a different cultural or temporal affiliation or whether the analysis results might be used to infer differences in the intensity and diversity of activities performed there.

In evaluating the potential for consistency in site interpretations, only the artifacts identified from each collection were considered, even though more artifacts, as well as features, have been documented at both sites. For the current analysis, the important consideration is whether site interpretations could differ based on differences in the results of in-field, digital-photograph, and hands-on laboratory analysis.

CHAPTER 3

Ceramic Analysis

In this chapter, the results of the four different analyses of ceramic artifacts performed at each of the two project sites are discussed. For the Fort Bliss site, the number of ceramic artifacts was limited (n = 20). However, many of these artifacts were of specific ceramic types rather than generic types, such as plain ware or buff ware. The number of ceramic artifacts analyzed for the Fort Huachuca site was much larger (n = 345). Although there were a large number of specific ceramic types in the collection, the majority of the artifacts were generic types, principally plain ware ceramic artifacts.

The focus of discussion is placed on variation among analyses in the identification of ceramic artifacts according to a series of four ceramic-artifact attributes: vessel element, ceramic ware, ceramic type, and the inferred temporal period associated with a particular ceramic type. For each of these attributes, several metrics introduced in Chapter 2 are calculated: agreement index, false-negative and false-positive rates, Cohen's kappa, and McNemar's test for bias. Agreement indexes were calculated for all pairs of analyses in order to provide an overall understanding of which analyses were more or less in agreement. The other metrics were calculated by assuming that hands-on laboratory analysis provided the most-accurate results, and it functioned as the "gold standard" in artifact identification for this project.

In addition to the above metrics, the distributions of attribute values were compared among analyses in order to evaluate the degree to which site interpretations would converge or diverge if site interpretation were to be based on the results of the in-field, digital-photograph, or hands-on laboratory analysis. Whereas the above metrics are based on per-artifact comparisons (either between pairs of analyses or between the hands-on analysis and the other analyses), the distributional comparisons investigated the potential implications that the different analyses could have on the interpretation of a site, regardless of whether the identification of an individual artifact was the same or different between analyses. What matters in these comparisons is whether the overall distribution of results is consistent or inconsistent among analyses.

Below, each of the ceramic attributes focused on for analysis (vessel element, ceramic ware, ceramic type, and temporal affiliation) are discussed in turn. This is followed by a discussion of the results from calculations of Cohen's kappa and McNemar's test for bias for select artifact attributes and then by a summary of the results.

Vessel Element

Vessel elements were recorded as either body sherds, rim sherds, or indeterminate. Rim sherds and body sherds were identified at both sites, and indeterminate vessel elements were identified only at Fort Huachuca. At both sites, body sherds were more numerous than rim sherds. However, at the Fort Bliss site, rim sherds were relatively more common, representing approximately one-third of recorded sherds. In contrast, less than 7 percent of sherds at the Fort Huachuca site were recorded as rim sherds.

Among all the ceramic attributes that were evaluated, vessel element was the most consistently recorded attribute among observers. The agreement index for vessel elements was higher than the agreement indexes of the other ceramic-artifact attributes, averaging 0.89 for Fort Bliss identifications and 0.93 for Fort Huachuca identifications (Tables 4 and 5). This result is promising, in that it suggests that the identification of a sherd as a body sherd or a rim sherd is relatively consistent among observers. Interestingly,

Table 4. The Agreement Index for Ceramic-Artifact Analysis at Fort Bliss, According to Vessel Element, Ceramic Ware, Ceramic Type, and Temporal Period

Observer 1	Observer 2	Vessel-Element Agreement	Ceramic-Ware Agreement	Ceramic-Type Agreement	Temporal-Period Agreement	Overall Agreement
Hands-on laboratory specialist	Field Technician 1	0.85	0.65	0.55	0.55	0.68
Hands-on laboratory specialist	Field Technician 2	0.95	0.75	0.60	0.60	0.77
Hands-on laboratory specialist	Digital-photograph laboratory specialist	0.95	0.70	0.55	0.69	0.73
Field Technician 1	Field Technician 2	0.80	0.85	0.75	0.78	0.80
Field Technician 1	Digital-photograph laboratory specialist	0.90	0.60	0.45	0.50	0.65
Field Technician 2	Digital-photograph laboratory specialist	0.90	0.70	0.50	0.53	0.70
Average		0.89	0.71	0.57	0.61	0.72
CV (%)		6.6	12.2	18.2	17.6	7.7

Key: CV = coefficient of variation.

Table 5. The Agreement Index for Ceramic-Artifact Analysis at Fort Huachuca, According to Vessel Element, Ceramic Ware, Ceramic Type, and Temporal Period

Observer 1	Observer 2	Vessel-Element Agreement	Ceramic-Ware Agreement	Ceramic-Type Agreement	Temporal-Period Agreement	Overall Agreement
Hands-on laboratory specialist	Field Technician 1	0.95	0.82	0.58	0.07	0.78
Hands-on laboratory specialist	Field Technician 2	0.94	0.67	0.46	0.09	0.69
Hands-on laboratory specialist	Digital-photograph laboratory specialist	0.92	0.76	0.49	0.27	0.72
Field Technician 1	Field Technician 2	0.97	0.76	0.48	0.54	0.74
Field Technician 1	Digital-photograph laboratory specialist	0.91	0.77	0.39	0.22	0.69
Field Technician 2	Digital-photograph laboratory specialist	0.89	0.62	0.50	0.17	0.67
Average		0.93	0.73	0.49	0.23	0.72
CV (%)		3.1	10.1	12.4	75.9	5.8

Key: CV = coefficient of variation.

the agreement index was highest between the two field technicians at the Fort Huachuca site but was lowest between the two field technicians at the Fort Bliss site, suggesting a relatively broad degree of variance between field observations for vessel element.

If we assume that the hands-on laboratory identification is most likely to be correct, then sherds at both sites were classified correctly according to vessel element for the vast majority of cases (Tables 6 and 7). Overall, the levels of Type I and Type II errors were low and were fairly similar among in-field and digital-photograph analyses, as indicated by the false-negative and false-positive rates, respectively. Field Technician 1 at Fort Bliss was slightly less accurate than the other analysts in identifying vessel element.

Rim sherds were evidently more difficult to identify than body sherds. At both Fort Bliss and Fort Huachuca, body sherds were rarely misidentified. Rim sherds were misclassified as body sherds or indeterminate vessel elements for nearly one-quarter of identifications made in the field or using only digital photographs. At Fort Bliss, no artifacts identified as body sherds by the hands-on analysis were misidentified as rim sherds. At Fort Huachuca, fewer than 5 percent of sherds identified as body sherds by the hands-on analysis were misidentified as rim sherds. These results suggest that close to one-quarter of rim sherds could go unnoticed when sherds are not examined by a specialist in a laboratory setting, but nearly all body sherds will be classified correctly in the field or using only digital photographs. At Fort Huachuca, sherds identified by the hands-on analysis as indeterminate were identified as rim or body sherds by other analyses; sherds identified as indeterminate by the in-field or digital-photograph analyses were identified as body or rim sherds by the hands-on laboratory analysis.

These results are important, because rim sherds are often considered to be potentially more informative than body sherds, in that they can convey important information about vessel-shape and -orifice characteristics. Rim sherds also are sometimes considered to more faithfully reflect vessel counts than body sherds, because rim sherds are typically far less numerous than body sherds. Rim sherds generally correspond to a much smaller percentage of a vessel than do body sherds, for instance.

Despite the generally high level of agreement in the identification of ceramic artifacts according to vessel element, difference among observers would potentially result in differences in the number of vessels inferred to have been present at a site. At Fort Bliss, three of the observers identified either six or seven rim sherds, resulting in a body-to-rim-sherd ratio of close to 2 to 1 (Table 8), but only four rim sherds were identified by Field Technician 1, resulting in a body-to-rim-sherd ratio roughly twice that identified by the other observers. Reliance on the in-field analysis performed by Field Technician 1 could theoretically have resulted in the interpretation of fewer vessels at the site, suggesting the possibility of a less-intensive occupation than what likely would have been inferred from the identifications made by other observers. There is a somewhat similar situation for the Fort Huachuca site, although differences among observers are not as great, considering the larger number of sherds. At Fort Huachuca, laboratory analysis also resulted in the identification of more rim sherds than did in-field analysis (Table 9). For both sites, the largest number of rim sherds was identified by the hands-on laboratory specialist, suggesting that rim sherds are best identified by a specialist in a controlled laboratory setting.

Wares

In the 1920s and 1930s, Harold S. Colton and Lyndon L. Hargrave (1937) developed a ceramic-classification system for prehistoric southwestern ceramics that is still used today. Colton defined a ware-series-type system. A ceramic series refers to a group of types discriminated on the basis of painted-design change and provides a subdivision within wares. Ceramic wares represent a group of ceramics similar in clays and tempers and manufacturing techniques. In a very gross sense, different wares often represent different cultural groups and have distinctive geographic distributions. In the context of this study, observations representing a diversity (or lack thereof) of ceramic wares could be indicative of interactions with different groups (or use by different

Table 6. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Vessel Element

Vessel Element	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Body sherd	—	—	—	—	0.23	0.08	—	0.10
Rim sherd	0.43	0.14	0.14	0.24	—	—	0.14	0.05
Total	0.15	0.05	0.05	0.08	0.15	0.05	0.05	0.08

Note: Cells shaded in gray indicate low rates, and the cell without shading indicates a moderate rate.

Table 7. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Vessel Element

Vessel Element	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Body sherd	0.03	0.04	0.04	0.04	0.03	0.02	0.03	0.02
Rim sherd	0.21	0.25	0.25	0.24	0.04	0.13	0.17	0.11
Indeterminate	**1.00**	0.50	**1.00**	**0.83**	**2.25**	**2.50**	**2.75**	**2.50**
Total	0.05	0.06	0.07	0.06	0.05	0.06	0.07	0.06

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 8. Artifact Counts at Fort Bliss, According to Analysis Type, per Vessel Element

Vessel Element	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Body sherd	16	14	13	14	57
Rim sherd	4	6	7	6	23
Total	20	20	20	20	80
Body-to-rim-sherd ratio	4	2.33	1.86	2.33	2.48

Table 9. Artifact Counts at Fort Huachuca, According to Analysis Type, per Vessel Element

Vessel Element	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Body sherd	316	312	317	312	1,257
Rim sherd	20	21	24	22	87
Indeterminate	9	12	4	11	36
Total	345	345	345	345	1,380
Body-to-rim-sherd ratio	15.80	14.86	13.21	14.18	14.45

groups) over a broad or limited geographic area and could impact some of the basic interpretations regarding what types of activities occurred at a site.

For the Fort Bliss site, in total, eight wares were identified among all observers (Table 10). The majority of sherds, regardless of observer, were identified as Jornada-region ceramics. A few wares were identified by only one observer: brown ware, Northern Rio Grande ware, and red ware. Most of these unique ware identifications were made in the laboratory, using digital photographs only, suggesting that information that could be derived from the digital photographs sometimes led to erroneous identifications of ceramic wares. This result ultimately led to an elevated level of ceramic-ware richness identified at the site, a result that could be interpreted, in the absence of other evidence, as indicative of either a more-diverse use of the site by multiple groups or a wider sphere of interaction for users of the site than what was suggested by the in-field or hands-on laboratory analysis of the same ceramic artifacts.

At the Fort Huachuca site, 12 wares were identified among observers, although the ware for some sherds was not determined (Table 11). The number of wares identified at the site is larger than what was found at the Fort Bliss site, but this is, in part, because of the much larger number of sherds identified at the Fort Huachuca site. In addition, the Fort Huachuca site is located near the peripheries of multiple culture areas and can be considered a kind of cultural crossroads or borderlands area where multiple cultural groups interacted. Most sherds at the Fort Huachuca site were identified as plain ware sherds. Hohokam Buff Ware and indeterminate painted wares were also commonly identified. Several wares were unique among observers: brown or buff ware, brown or red ware, buff ware, San Simon series, Trincheras, and Upper/Middle Santa Cruz River valley tradition. Interestingly, three of these unique identifications were made during the digital-photograph analysis, and all of them were relatively general ware identifications (brown or buff ware, brown or red ware, or buff ware). This result suggests that the digital photographs were inadequate in a number of cases to make more-specific ware identifications, in comparison to a hands-on analysis. In addition, the identification of red ware sherds was only common for the digital-photograph analysis, suggesting that digital photographs were inadequate to differentiate between red wares and other wares, a problem that could potentially have resulted from the color balance of the digital photographs. It is worth noting, also, that the single identification of a red ware sherd at the Fort Bliss site was also made during the digital-photograph analysis.

On average, the agreement index was 0.71 for Fort Bliss ware identifications and 0.73 for Fort Huachuca ware identifications (see Tables 4 and 5). In other words, slightly less than three-quarters of ware identifications were in agreement with each other, on average. The highest agreement index was calculated for ware identifications made by field technicians at Fort Bliss. The agreement index between identifications made by field technicians at Fort Huachuca was closer to the average for ware identifications at the site.

For both Fort Bliss and Fort Huachuca, overall error rates were low or moderate, because the most-common wares at both sites were correctly identified most of the time (Jornada-region ceramics at Fort Bliss and plain ware at Fort Huachuca) (Tables 12 and 13). False-negative rates, however, were high for all other wares, indicating that the rarer wares were rarely identified correctly. False-positive rates at both sites tended to be low or moderate, because most wares present in the collections were rarely

Table 10. Artifact Counts at Fort Bliss, According to Analysis Type, per Ceramic Ware

Ceramic Ware	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Brown ware	—	1	—	—	1
Chihuahua tradition	—	—	1	1	2
Indeterminate painted ware	2	—	1	1	4
Jornada-region ceramics	17	19	15	14	65
Mimbres Mogollon ceramics	—	—	—	2	2
Northern Rio Grande ware	—	—	—	1	1
Plain ware	1	—	3	—	4
Red ware	—	—	—	1	1
Total	20	20	20	20	80
Richness	3	2	4	6	8

Table 11. Artifact Counts at Fort Huachuca, According to Analysis Type, per Ceramic Ware

Ceramic Ware	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Brown or buff ware	—	—	—	1	1
Brown or red ware	—	—	—	2	2
Buff ware	—	—	—	11	11
Hohokam Buff Ware	37	34	11	16	98
Indeterminate painted ware	4	62	7	7	80
Plain ware	288	237	290	278	1,093
Red ware	—	—	1	15	16
San Pedro River valley tradition	1	—	10	1	12
San Simon series	—	—	4	—	4
Trincheras	2	—	—	—	2
Tucson Basin ceramics	4	—	19	8	31
Upper/Middle Santa Cruz River valley tradition	—	—	1	—	1
Not determined	9	12	2	6	29
Total	345	345	345	345	1,380
Richness	6	3	8	9	12

Table 12. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Ceramic Ware

Ceramic Ware	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Chihuahua tradition	**1.00**	**1.00**	—	**0.67**	—	—	—	—
Indeterminate painted ware	**1.00**	**1.00**	**1.00**	**1.00**	**2.00**	—	**1.00**	**1.00**
Jornada-region ceramics	0.13	—	0.13	0.09	0.27	0.27	0.07	0.20
Plain ware	**1.00**	**1.00**	**1.00**	**1.00**	0.33	—	—	0.11
Total	0.35	0.25	0.30	0.30	0.35	0.20	0.10	0.22

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 13. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Ceramic Ware

Ceramic Ware	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Hohokam Buff Ware	**0.55**	**0.55**	**0.55**	**0.55**	**2.91**	**2.64**	**1.00**	**2.18**
Indeterminate painted ware	**0.86**	**1.00**	**0.86**	**0.90**	0.43	**8.86**	**0.86**	**3.38**
Plain ware	0.04	0.22	0.12	0.13	0.03	0.03	0.08	0.05
Red ware	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**15.00**	**5.00**
San Pedro River valley tradition	**1.00**	**1.00**	**0.90**	**0.97**	0.10	—	—	0.03
San Simon series	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Tucson Basin ceramics	**1.00**	**1.00**	**0.74**	**0.91**	0.21	—	0.16	0.12
Upper/Middle Santa Cruz River valley tradition	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Total	0.17	0.32	0.22	0.24	0.14	0.29	0.17	0.20

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

identified by in-field or digital-photograph analysis. False-positive rates, however, were quite high for indeterminate painted ware at Fort Bliss. At Fort Huachuca, false-positive rates were high for Hohokam Buff Ware in all analyses and were also high in one or more analyses for indeterminate painted wares and red ware.

An interesting result of the Fort Huachuca ceramic analyses is that one of the field technicians who made identifications at the Fort Huachuca site identified a much larger number of sherds as indeterminate painted wares than did any other analyst. This accounts for much of the difference in misidentification rates for ceramic wares between in-field analyses. Although most of the analysts identified approximately 2 percent of sherds as indeterminate painted ware sherds, one field analyst identified nearly 18 percent of sherds as indeterminate painted ware. At the same time, although most of the analysts identified approximately 80 percent of sherds as plain ware, the same field analyst identified less than 70 percent of sherds as plain ware. These results suggest that the field technician potentially had less familiarity with ceramic traditions found in the region and also had a greater tendency to identify plain wares as painted wares. If interpretations were to rely on only the results of this one in-field analysis, plain ware sherds would be interpreted as less common than their likely occurrence. Moreover, the kinds of painted wares present at the site would be substantially less certain, a result that could potentially require additional and costly further visitation and documentation of the site in order to seek additional information or evaluate site significance.

A particularly interesting result of the Fort Huachuca ware identifications is that field technicians tended to identify sherds as Hohokam Buff Ware substantially more often than did trained specialists (see Table 11). Indeed, the fewest Hohokam Buff Ware identifications were made during the hands-on laboratory identifications. In contrast, the greatest number of Tucson Basin ceramic ware identifications, by far, was made during the hands-on laboratory analysis. This result is particularly important in terms of interpreting the potential interaction sphere of the Fort Huachuca site. With in-field analysis only, and in the absence of other evidence for cultural affiliation, an investigator would likely interpret the site's inhabitants as interacting most intensively with the Gila-Salt Basin Hohokam or perhaps even infer that the site was occupied mostly by people affiliated with the Gila-Salt Basin Hohokam. Interaction with the Tucson Basin Hohokam would be interpreted as absent or minimal. Based on the hands-on laboratory analysis, an investigator would, instead, interpret interactions with both the Gila-Salt Basin Hohokam and the Tucson Basin Hohokam to have occurred and may interpret interactions with the Tucson Basin Hohokam to have been more intensive than interactions with the Gila-Salt Basin Hohokam.

In this particular site context, such an interpretation could also have influenced the interpretation of how the site related behaviorally to other sites in the region. As a result of the in-field analysis, the site might be interpreted as more closely connected to sites to the north, along the lower San Pedro and Lower Gila Rivers, because sites along these drainages more often include ceramic artifacts associated with the Gila-Salt Basin Hohokam and are often interpreted as more closely connected geographically and culturally with sites in the Gila-Salt Basin. In contrast, as a result of the hands-on laboratory analysis, the site may be interpreted as more closely connected to sites located in the Tucson Basin, along the Santa Cruz River, as well as to intervening sites located along Cienega Creek, between the Fort Huachuca site and the Tucson Basin. In other words, interpretations of who the site inhabitants were, in terms of cultural affiliation and with which groups they interacted, could be quite different, depending on which results were used to interpret the site, with a marked difference between hands-on laboratory analysis and other analyses.

Several ware identifications made during the hands-on laboratory analysis of the Fort Huachuca site were unique to the hands-on laboratory analysis, including San Simon–series and Upper/Middle Santa Cruz River valley–tradition ware identifications. In addition, the identification of a San Pedro River valley–tradition ware was substantially more common for the hands-on laboratory identifications and was rare for or absent from identifications made during in-field or digital-photograph analysis. The San Simon–series, Upper/Middle Santa Cruz River valley–tradition, and San Pedro River valley–tradition wares are all associated with more-local and geographically distinctive ceramic traditions than are the Hohokam-affiliated wares. Interactions with the makers of ceramic vessels associated with the more-local traditions (or use of the site by the makers of such ceramics) could only be inferred from the results of the hands-on laboratory analysis.

Ceramic Types

In total, 18 and 29 different ceramic types were identified by different analysts at the Fort Bliss (Table 14) and Fort Huachuca (Table 15) sites, respectively. For both sites, more types were identified by trained specialists in a laboratory setting than were identified by field technicians in the field. This is not particularly surprising, because specialists would likely have greater knowledge and familiarity with a wider array of ceramic types, as well as a greater level of confidence in assigning a sherd to a particular ceramic type. In the case of Fort Huachuca, on the order of twice as many ceramic types were identified by trained specialists in a laboratory setting than by the field technicians.

The agreement index for ceramic types at the Fort Bliss site was 0.57, and it was 0.49 for ceramic types at the Fort Huachuca site, on average (see Tables 4 and 5). The lower agreement index for ceramic types than for ceramic wares is understandable, given the larger number of ceramic types identified and the greater level of expertise required to identify ceramics according to type. Overall, this result indicates that, on average, slightly more than one-half of the identifications of ceramic type were in agreement between sets of observations.

Several types were identified during the hands-on laboratory analysis for both sites that were not identified by either in-field analysis or digital-photograph analysis. Similarly, some types identified by in-field analysis or digital-photograph analysis were not identified during the hands-on laboratory analysis. At Fort Bliss, El Paso Bichrome (red-on-brown), indeterminate fine paste (unpainted), and Type II (smooth, sand) ceramic artifacts were identified only during the hands-on laboratory analysis. Brown corrugated, El Paso Bichrome, Galisteo Black-on-white, indeterminate polychrome, red ware, red ware with El Paso paste (red slip), San Andres Red-on-terracotta, San Francisco Red, and an unidentified textured type were not identified during the hands-on laboratory analysis but were identified during at least one other analysis of the same materials. It appears that at Fort Bliss, specific types that may not actually have been present in the collection were identified by in-field or digital-photograph analysis. This is particularly the case for the digital-photograph analysis, which identified six ceramic types that were not identified by any other analysis.

At Fort Huachuca, Babocomari Bichrome, Casa Grande Red-on-buff, Encinas Red-on-brown, indeterminate San Simon Red-on-brown, Rincon or Tanque Verde Red-on-brown, and Santa Cruz Polychrome were identified only during the hands-on laboratory analysis. Brown or buff ware, indeterminate red-on-brown ware (elaborated), indeterminate black-on-gray ware, indeterminate brown or red ware, indeterminate buff ware, indeterminate plain ware, indeterminate red or brown plain ware, Tanque Verde Red-on-brown, Trincheras Purple-on-red (nonspecular), and Type IV (phyllite inclusions) were not identified during the hands-on laboratory analysis but were identified during at least one other analysis of the same materials. It appears that, in place of specific types identified by the hands-on analysis, indeterminate types were identified by other analyses.

Another interesting difference among analyses of the Fort Huachuca ceramic artifacts is that the digital-photograph analysis identified approximately 85 percent of ceramic artifacts as Type II (smooth, sand), whereas for the other analyses, closer to 50 percent of ceramic artifacts were identified as Type II (smooth, sand). At the same time, most analyses identified a substantial percentage of other ceramic artifacts as Type III (micaceous schist), whereas no such artifacts were identified by the digital-photograph analysis. This suggests that digital photography was inadequate for differentiating between sand and micaceous-schist inclusions, a result that could have a substantial bearing on the interpretation of ceramic technology at the site, as well as interpretations of ceramic manufacture and provenance, because pastes bearing either sand or micaceous schist theoretically could have different performance characteristics and could be derived from different clay or temper sources.

At Fort Bliss, the correspondence of identified types is relatively close between the in-field analysis and the hands-on laboratory analysis and is most disparate between the digital-photograph analysis and the other analyses. This suggests that, on a general level, the in-field analysis did a fair job at identifying ceramics according to type at Fort Bliss. The digital-photograph analysis, in contrast, resulted in the identification of a

Table 14. Artifact Counts at Fort Bliss, According to Analysis Type, per Ceramic Type

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Brown corrugated	—	1	—	—	1
Chupadero Black-on-white	4	4	4	3	15
El Paso Bichrome	2	2	—	—	4
El Paso Bichrome (red-on-brown)	—	—	1	—	1
El Paso Brown	2	2	2	2	8
El Paso Polychrome	2	2	4	2	10
Galisteo Black-on-white	—	—	—	1	1
Indeterminate fine paste (unpainted)	—	—	2	—	2
Indeterminate polychrome	—	—	—	1	1
Indeterminate red-on-brown ware	2	—	1	—	3
Playas Red Incised	—	—	1	1	2
Red ware	—	—	—	1	1
Red ware with El Paso paste (red slip)	—	—	—	2	2
San Andres Red-on-terracotta	—	—	—	1	1
San Francisco Red	—	—	—	2	2
Textured	1	—	—	—	1
Three Rivers Red-on-terracotta	7	9	4	4	24
Type II (smooth, sand)	—	—	1	—	1
Total	20	20	20	20	80
Richness	7	6	9	11	18

Table 15. Artifact Counts at Fort Huachuca, According to Analysis Type, per Ceramic Type

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Babocomari Bichrome	—	—	1	—	1
Brown or buff ware	—	—	—	1	1
Casa Grande Red-on-buff	—	—	1	—	1
Dragoon Red-on-brown (elaborated)	1	—	9	1	11
Encinas Red-on-brown	—	—	1	—	1
Indeterminate red-on-brown ware (broad line)	—	—	1	4	5
Indeterminate red-on-brown ware (elaborated)	—	—	—	1	1
Indeterminate red-on-brown ware (fine line)	—	—	1	1	2
Indeterminate Tucson Basin Black-on-brown	4	—	1	—	5

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Indeterminate Tucson Basin Red-on-brown	—	—	10	3	13
Indeterminate black-on-gray ware	—	1	—	—	1
Indeterminate brown or red ware	—	—	—	1	1
Indeterminate buff ware (no paint)	—	—	3	2	5
Indeterminate buff ware	—	—	—	11	11
Indeterminate painted ware	1	61	2	—	64
Indeterminate plain ware	—	—	—	1	1
Indeterminate red or brown plain ware	—	—	—	1	1
Indeterminate red-on-brown ware	3	—	5	1	9
Indeterminate red-on-buff ware	37	34	7	14	92
Indeterminate San Simon Red-on-brown	—	—	3	—	3
Red Type II (smooth, sand)	—	—	1	15	16
Rincon or Tanque Verde Red-on-brown	—	—	1	—	1
Rincon Red-on-brown	—	—	5	3	8
Santa Cruz Polychrome	—	—	1	—	1
Tanque Verde Red-on-brown	—	—	—	2	2
Trincheras Purple-on-red (nonspecular)	2	—	—	—	2
Type II (smooth, sand)	152	196	188	277	813
Type III (micaceous schist)	130	40	102	—	272
Type IV (phyllite)	6	1	—	—	7
Not determined	9	12	2	6	29
Total	345	345	345	345	1,380
Richness	9	6	19	17	29

number of additional specific ceramic types not identified by other analyses. If digital photographs alone had been used to identify ceramic artifacts according to type at the site, the interpretation of the site's ceramic assemblage would likely have been substantially different from an interpretation based on the results of either the in-field or the hands-on laboratory analysis.

At Fort Huachuca, multiple specific ceramic types were identified by the hands-on laboratory analysis that were not identified by the other analyses, including several that are associated with local ceramic traditions not identified by the other analyses. As indicated above, ceramic types that were rarely identified by the hands-on laboratory analysis but were more commonly identified by other analyses were sometimes less-specific types (e.g., indeterminate red-on-buff or indeterminate painted ware). In addition, ceramic types identified as indeterminate red-on-buff by the in-field or digital-photograph analysis were often identified as one of several types of red-on-brown ceramic types, such as indeterminate Tucson Red-on-brown, Rincon Red-on-brown, Dragoon Red-on-brown (elaborated), or Encinas Red-on-brown. In other words, differentiating buff wares from brown wares was difficult during the in-field and digital-photograph analyses. This result suggests that more-detailed, more-accurate, and more-precise results were achieved by the hands-on laboratory analysis than by the in-field or digital-photograph analysis.

Calculation of false-positive and false-negative rates suggests that, at Fort Bliss, low rates were calculated only for Chupadero Black-on-white and El Paso Brown (Table 16). Low false-negative rates were also calculated for Three Rivers Red-on-terracotta, but false-positive rates were high for this ceramic type for in-field analysis, indicating that the type was frequently misapplied to other ceramic types by the field technicians. El Paso Polychrome had a moderate false-negative rate and a low false-positive rate, indicating that analysts performed fairly well in identifying the type. Other ceramic types mostly had high false-negative rates and were misidentified 100 percent of the time by several analysts. With the exception of indeterminate red-on-brown ware, which was identified erroneously in two cases by Field Technician 1, these ceramic types also had low false-positive rates, because they were not identified by the in-field or digital-photograph analyses.

The results for the Fort Huachuca site are similar; most ceramic types have high false-negative rates and low false-positive rates, because they were either not identified or only rarely identified by the in-field or digital-photograph analysis (Table 17). Overall error rates from the analyses of the Fort Huachuca site were also similar to but slightly higher than those from the Fort Bliss site. As with ceramic wares at Fort Huachuca, error rates were highest for the analysis performed by Field Technician 2 and lowest for the analysis performed by Field Technician 1. The difference in error rates between in-field analyses stems mostly from the fact that Field Technician 2 misidentified the relatively common Type III (micaceous schist) sherds in over two-thirds of cases and often identified sherds erroneously as indeterminate painted ware or indeterminate red-on-buff sherds. Field Technician 1 misidentified Type III (micaceous-schist inclusions) for approximately one-quarter of cases but also applied the label incorrectly to other types. Otherwise, misidentification rates are high and very similar between the two in-field analyses for ceramic types.

Overall, sherds of only two ceramic types were misidentified for a minority of cases at Fort Huachuca. Type II (smooth, sand) sherds were misidentified for approximately one-quarter of cases, as indicated by the false-negative rates; indeterminate red-on-buff sherds were misidentified for around 40 percent of cases, as indicated by the false-negative rates. The digital-photograph analysis misidentified Type II (smooth, sand) sherds substantially less often than did the in-field analysis. At the same time, however, false-positive rates show that the label "indeterminate red-on-buff" was frequently misapplied by all analysts, as was "Type II (smooth, sand)" by the digital-photograph analysis. The digital-photograph analysis also frequently misapplied the label "Red Type II (smooth, sand inclusions)" to sherds of other types and, more rarely, misapplied the labels "indeterminate red-on-brown ware (broad line)" and "indeterminate red-on-brown ware (fine line)" to ceramic artifacts of other types.

All together, these results suggest that many sherds, particularly those corresponding to specific types that can be used to more precisely determine ceramic technology or temporal or cultural affiliation, would be misidentified, were they only to be subjected to in-field analysis by field technicians or to digital-photograph analysis by a trained specialist. These results further suggest that interpretations of the site based on ceramic types would differ substantially according to the results of the in-field, digital-photograph, and hands-on laboratory analyses. The relative proportions of different ceramic types at the site (which are sometimes used to interpret a site's temporal or cultural affiliation) differ substantially among analyses, for instance. The most nuanced and detailed interpretation of the site in terms of chronology, cultural affiliation, and ceramic technology would be achieved as a result of the hands-on analysis. Thus, it appears that the most useful information potentially contributing to evaluation of a site would be information derived from hands-on laboratory identification by a trained specialist, not information obtained on ceramic types by the other types of analyses.

Table 16. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Ceramic Type

Ceramic Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Chupadero Black-on-white	—	—	0.25	0.08	—	—	—	—
El Paso Bichrome (red-on-brown)	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
El Paso Brown	—	—	—	—	—	—	—	—
El Paso Polychrome	0.50	0.50	0.50	0.50	—	—	—	—
Indeterminate fine paste (unpainted)	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Indeterminate red-on-brown ware	**1.00**	**1.00**	**1.00**	**1.00**	**2.00**	—	—	**0.67**
Playas Red Incised	**1.00**	**1.00**	—	**0.67**	—	—	—	—
Three Rivers Red-on-terracotta	0.25	—	0.25	0.17	**1.00**	**1.25**	0.25	**0.83**
Type II (smooth, sand)	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Total	0.45	0.40	0.45	0.43	0.30	0.25	0.05	0.20

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 17. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Ceramic Type

Ceramic Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Babocomari Bichrome	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Casa Grande Red-on-buff	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Dragoon Red-on-brown (elaborated)	**1.00**	**1.00**	**0.89**	**0.96**	0.11	—	—	0.04

continued on next page

Ceramic Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	**Field Technician 2**	**Digital-Photograph Laboratory Specialist**	**Total**	**Field Technician 1**	**Field Technician 2**	**Digital-Photograph Laboratory Specialist**	**Total**
Encinas Red-on-brown	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Indeterminate red-on-brown ware (broad line)	**1.00**	**1.00**	—	**0.67**	—	—	**3.00**	**1.00**
Indeterminate red-on-brown ware (fine line)	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**1.00**	0.33
Indeterminate Tucson Basin Black-on-brown	**1.00**	**1.00**	**1.00**	**1.00**	**4.00**	—	—	**1.33**
Indeterminate Tucson Basin Red-on-brown	**1.00**	**1.00**	**1.00**	**1.00**	—	—	0.30	0.10
Indeterminate buff ware (no paint)	**1.00**	**1.00**	**0.67**	**0.89**	—	—	0.33	0.11
Indeterminate painted ware	**1.00**	**1.00**	**1.00**	**1.00**	0.50	**30.50**	—	**10.33**
Indeterminate red-on-brown ware	**1.00**	**1.00**	**1.00**	**1.00**	**0.60**	—	0.20	0.27
Indeterminate red-on-buff ware	0.43	0.43	0.43	0.43	**4.71**	**4.29**	**1.43**	**3.48**
Indeterminate San Simon Red-on-brown	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Red Type II (smooth, sand)	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**15.00**	**5.00**
Rincon or Tanque Verde Red-on-brown	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Rincon Red-on-brown	**1.00**	**1.00**	0.40	**0.80**	—	—	—	—
Santa Cruz Polychrome	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Type II (smooth, sand)	0.34	0.34	0.13	0.27	0.15	0.38	**0.61**	0.38
Type III (mica-ceous schist)	0.25	**0.70**	**1.00**	**0.65**	**0.52**	0.09	—	0.20
Total	0.40	**0.53**	0.50	0.48	0.36	0.50	0.43	0.43

Note: Cells shaded in gray indicate low rates and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Estimates of the Minimum Number of Vessels

A large number of sherds could correspond to one vessel or many, depending on ceramic type, vessel size, and degree of vessel fragmentation, as well as other formation processes. Rather than individual sherd counts, an estimate of the minimum number of vessels can better approximate the intensity of ceramic-artifact use. Ceramic-vessel-element and ceramic-type data were used to estimate the minimum number of vessels present at each site. To calculate the minimum number of vessels, each rim sherd for a particular ceramic type was counted, per analysis, as one vessel. If no rim sherds were identified by an analysis for a particular ceramic type, but at least one body sherd of that type was identified, then a vessel count of one was assigned to the type for the analysis. If one or more rim sherds were identified for a particular ceramic type, then any additional body sherds of the ceramic types did not add to the vessel count, under the assumption that they could have come from a vessel for which a rim sherd had already been identified. It is worth noting that information on the refitting of ceramic artifacts could conceivably be used to refine vessel counts, but the above approach attempts to mimic the situation in which such information may not be available.

The results are instructive and complement the discussion above in regard to vessel elements and the number of rim sherds identified by the various analyses (Tables 18 and 19). For both Fort Bliss and Fort Huachuca, laboratory analyses resulted in vessel counts that were approximately 50 percent larger than vessel counts arrived at using in-field-analysis results. The vessel counts calculated using the digital-photograph-analysis results, although similar to the counts derived from the results of the hands-on analysis, were slightly larger or smaller at either site than the vessel counts calculated using the hands-on laboratory analysis. For either site, estimates of ceramic-use intensity based on the minimum number of vessels would be substantially underestimated if it were based on the in-field analysis. In contrast, vessel counts arrived at based on digital-photograph analysis would probably be largely accurate.

Inferred Temporal Period

A final comparison made between analyses in terms of ceramic-artifact identifications was in regard to temporal periods that could be inferred from the ceramic-type identifications. Many ceramic artifacts could not be assigned to a specific temporal period, particularly in cases in which large numbers of ceramic artifacts were plain ware artifacts that could not be assigned to more-specific types. Most of these ceramic artifacts should likely date to the period between A.D. 200 and 1450, when ceramic vessels are documented to have been manufactured during prehistory in the region. However, it is at least theoretically plausible that some could have been manufactured prior to A.D. 200 or after A.D. 1450, during the protohistoric or historical periods.

For the Fort Bliss site, on the order of three-quarters of the ceramic artifacts were identified to types that could be dated to specific temporal periods (Table 20). Most of these types are associated with dates spanning from A.D. 1125 to 1450. However, the hands-on laboratory analysis identified one type that dates to an earlier period, A.D. 900–1150 (El Paso Bichrome [red-on-brown]), and both laboratory analyses identified a type that dates to between A.D. 1200 and 1400 (Playas Red Incised). All analyses identified two specimens as El Paso Brown, a relatively generic ceramic-artifact type that dates to A.D. 200–1250. At Fort Bliss, overall patterns that could be inferred from ceramic-artifact counts, according to period, are relatively similar. In other words, most sherds with defined temporal affiliations date to the Late Formative period. As noted above, however, ceramic-vessel use during the Middle Formative period (between A.D. 900 and 1150) could only be inferred on the basis of the hands-on laboratory analysis. Thus, the hands-on laboratory ceramic-artifact analysis identified a temporal component that was not identified by the other ceramic-artifact analyses conducted in the field or using only digital photographs.

For the Fort Huachuca site, typically only around 10 percent of ceramic artifacts could be dated to a specific temporal period (Table 21). This is because most sherds lacked decoration, such as painted

Table 18. Estimated Minimum Number of Vessels at Fort Bliss, According to Analysis Type, per Ceramic Type

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Brown corrugated	—	1	—	—	1
Chupadero Black-on-white	1	1	1	1	4
El Paso Bichrome	1	1	—	—	2
El Paso Bichrome (red-on-brown)	—	—	1	—	1
El Paso Brown	2	2	2	2	8
El Paso Polychrome	1	2	3	1	7
Galisteo Black-on-white	—	—	—	1	1
Indeterminate fine paste (unpainted)	—	—	1	—	1
Indeterminate polychrome	—	—	—	1	1
Indeterminate red-on-brown ware	1	—	1	—	2
Playas Red Incised	—	—	1	1	2
Red ware	—	—	—	1	1
Red ware with El Paso paste (red slip)	—	—	—	1	1
San Andres Red-on-terracotta	—	—	—	1	1
San Francisco Red	—	—	—	1	1
Textured	1	—	—	—	1
Three Rivers Red-on-terracotta	1	1	1	2	5
Type II (smooth, sand)	—	—	1	—	1
Total	8	8	12	13	41

Table 19. Estimated Minimum Number of Vessels at Fort Huachuca, According to Analysis Type, per Ceramic Type

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Babocomari Bichrome	—	—	1	—	1
Brown or buff ware	—	—	—	1	1
Casa Grande Red-on-buff	—	—	1	—	1
Dragoon Red-on-brown (elaborated)	1	—	2	1	4
Encinas Red-on-brown	—	—	1	—	1
Indeterminate red-on-brown ware (broad line)	—	—	1	1	2
Indeterminate red-on-brown ware (elaborated)	—	—	—	1	1
Indeterminate red-on-brown ware (fine line)	—	—	1	1	2
Indeterminate Tucson Basin Black-on-brown	1	—	1	—	2
Indeterminate Tucson Basin Red-on-brown	—	—	2	1	3

Ceramic Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Indeterminate black-on-gray ware	—	1	—	—	1
Indeterminate brown or red ware	—	—	—	1	1
Indeterminate buff ware (no paint)	—	—	1	1	2
Indeterminate buff ware	—	—	—	1	1
Indeterminate painted ware	1	4	1	—	6
Indeterminate plain ware	—	—	—	1	1
Indeterminate red or brown plain ware	—	—	—	1	1
Indeterminate red-on-brown ware	1	—	2	1	4
Indeterminate red-on-buff ware	6	6	1	3	16
Indeterminate San Simon Red-on-brown	—	—	1	—	1
Red Type II (smooth, sand)	—	—	1	1	2
Rincon or Tanque Verde Red-on-brown	—	—	1	—	1
Rincon Red-on-brown	—	—	1	1	2
Santa Cruz Polychrome	—	—	1	—	1
Tanque Verde Red-on-brown	—	—	—	1	1
Trincheras Purple-on-red (nonspecular)	1	—	—	—	1
Type II (smooth, sand)	5	10	9	14	38
Type III (micaceous schist)	5	1	6	—	12
Type IV (phyllite)	2	1	—	—	3
Total	23	23	35	32	113

Table 20. Artifact and Vessel Counts at Fort Bliss, According to Analysis Type, per Temporal Period

Temporal Period (A.D.)	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
200–1250	2 (2)	2 (2)	2 (2)	2 (2)	8 (8)
900–1150	—	—	1 (1)	—	1 (1)
1125–1300	7 (1)	9 (1)	4 (1)	5 (3)	25 (6)
1150–1450	6 (2)	6 (3)	8 (4)	5 (2)	25 (11)
1200–1400	—	—	1 (1)	1 (1)	2 (2)
Not determined	5 (3)	3 (2)	4 (3)	7 (5)	19 (13)
Total	20 (8)	20 (8)	20 (12)	20 (13)	80 (41)

Note: Estimated vessel counts are in parentheses.

Table 21. Artifact and Vessel Counts at Fort Huachuca, According to Analysis Type, per Temporal Period

Temporal Period (A.D.)	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
650–750	—	—	1 (1)	4 (1)	5 (2)
700–1150	2 (1)	—	—	—	2 (1)
700–1300	37 (6)	34 (6)	11 (3)	16 (4)	98 (19)
750–950	—	—	1 (1)	1 (1)	2 (2)
950–1150	1 (1)	—	15 (4)	5 (3)	21 (8)
950–1300	—	—	1 (1)	—	1 (1)
1150–1300	—	—	1 (1)	2 (1)	3 (2)
Not determined	305 (15)	311 (17)	315 (24)	317 (22)	1,248 (78)
Total	345 (23)	345 (23)	345 (35)	345 (32)	1,380 (113)

Note: Estimated vessel counts are in parentheses.

design elements, that would allow their placement into types affiliated with specific temporal periods. For the in-field analysis at Fort Huachuca, most sherds that could be affiliated with a specific temporal period were associated with a relatively broad period spanning the Middle and Late Formative periods, from A.D. 700 to 1300. Field Technician 1, however, did identify a few sherds as dating to the Middle Formative period, one of which was dated to the period from A.D. 950 to 1150; the other two were dated more broadly to the period from A.D. 700 to 1150. Substantial numbers of sherds dating to the Middle Formative period were identified only by the laboratory analyses; most were identified as dating to the period from A.D. 950 to 1150, and one per laboratory-analysis type dates to the period from A.D. 750 to 950. In addition, both the hands-on and digital-photograph laboratory analyses identified at least one ceramic artifact dating to the transition between the Early and Middle Formative periods (A.D. 650–750) as well as to the early portion of the Late Formative period (A.D. 1150–1300).

Obviously, the temporal information derived from the laboratory analyses at Fort Huachuca was substantially more detailed and precise than the temporal information derived from the in-field analysis. Three discrete periods were missed completely by the in-field analyses (A.D. 650–750, 750–950, and 1150–1300). Given that the dating of ceramic artifacts is often the principal evidence used to infer temporal affiliation of Formative period sites in Arizona and New Mexico, the temporal periods missed by the in-field analysis would have likely affected site interpretation. The in-field analysis performed by Field Technician 1 identified sherds dating to the period from A.D. 700 to 1300 and a few dating to A.D. 950–1150; the in-field analysis performed by Field Technician 2 indicated only that site component(s) dated to sometime within the larger range of A.D. 700–1300.

Overall, the level of agreement among analyses for temporal affiliation at the Fort Bliss site was similar to the level of agreement among analyses for ceramic type (see Table 4). This appears to be because most ceramic artifacts at the Fort Bliss site were identified according to types that could be placed into a small number of temporal periods. At Fort Huachuca, the level of agreement was quite low; roughly one-quarter of observations were in agreement, on average (see Table 5). This is because in a substantial number of cases, one observer identified a ceramic artifact as a type that could be dated, and another observer identified the same ceramic artifact as a type that does not have a defined temporal affiliation. The result was a level of agreement regarding temporal affiliation that was, overall, much lower at the Fort Huachuca site than at the Fort Bliss site.

Interestingly, the highest level of agreement was calculated for both sites for temporal affiliations derived from the in-field analyses. For Fort Huachuca, this is largely because these analyses tended to identify most sherds according to ceramic types that could only be placed in the same broad period (i.e., A.D. 700–1300). However, the greatest level of agreement between the hands-on laboratory analysis and the other analyses was obtained, at both Fort Bliss and Fort Huachuca, between the hands-on laboratory analysis and the digital-photograph analysis. The level of agreement between the hands-on laboratory analysis and the in-field analysis was lower at both sites. Thus, even though hands-on laboratory analysis and digital-photograph analysis often disagreed on the identification of the ware and type of a ceramic artifact, the identifications made in a laboratory setting by trained specialists still identified the artifact as associated with the same time period more often than did the in-field ceramic-artifact identifications.

Overall, false-negative rates and false-positive rates for temporal period were highest for in-field analyses and lowest for digital-photograph analysis. At Fort Bliss, false-negative rates were low or moderate for all analyses for the periods A.D. 200–1250, 1125–1300, and 1150–1450 (Table 22). False-positive rates were high for the in-field analyses for the period A.D. 1125–1300, however, indicating that artifacts that are not associated with the period A.D. 1125–1300 were erroneously identified by in-field analysis as associated with that period. The few artifacts associated with the periods A.D. 900–1150 and 1200–1400 were misidentified by the in-field analysis in all cases, as were sherds with undetermined temporal affiliations. Sherds associated with the periods A.D. 900–1150 and 1200–1400 were not identified by the in-field analysis, resulting in low false-positive rates. Artifacts not affiliated with a defined temporal period were, however, erroneously identified by the in-field analysis, resulting in high false-positive rates. The digital-photograph analysis had low error rates for the period A.D. 1200–1400 and for sherds with undetermined temporal affiliations, but it erroneously identified a sherd as a type associated with the period A.D. 900–1150, resulting in a high false-positive rate for the period.

In many cases at Fort Bliss, the misidentification of a temporal period would potentially be of little consequence, because most of the temporal periods identified substantially overlap. However, the failure of the in-field analysis to identify a component dating to the period A.D. 900–1150 and the failure of the digital-photograph analysis to correctly identify a sherd with that period would have resulted in a failure to identify a discrete temporal component at the site, were these analyses relied upon to interpret site chronology. The only other ceramic artifacts spanning the Early and Middle Formative periods date to a much broader span, from A.D. 200–1250, and could date to any time during the Early or Middle Formative period or to sometime during the early portion of the Late Formative period.

At Fort Huachuca, overall error rates for temporal period were quite low for both the in-field and the digital-photograph analyses (Table 23). This is because most ceramic artifacts in the collection from the sites were types that could not be affiliated with temporal periods. Moreover, these types were rarely incorrectly identified by the in-field or digital-photograph analyses as types that could be affiliated with a defined temporal period. However, nearly all defined temporal-period affiliations were misidentified for most or all cases by the in-field or digital-photograph analysis. False-negative rates were high for all periods identified, with exception of the period A.D. 650–750 for the digital-photograph analysis. The temporal affiliation of the one artifact dating to that period was correctly identified only by the digital-photograph analysis. False-positive rates were low for most periods for in-field analysis, because it almost never identified artifacts associated with most of the periods that were identified by the hands-on laboratory analysis. In contrast, false-positive rates were high for several periods for the digital-photograph analysis, because although artifacts affiliated with these periods were identified by the digital-photograph analysis, they were identified erroneously.

The misidentification of ceramic types at Fort Huachuca would potentially have substantial consequences for the interpretation of site chronology, particularly for in-field analysis. In-field analysis tended to identify artifacts according to types affiliated with broad periods, whereas the hands-on laboratory and digital-photograph analyses more often identified ceramic types affiliated with shorter and more-precisely defined temporal periods. Intriguingly, although ceramic-type identifications resulting from the hands-on laboratory and digital-photograph analyses would have resulted in the identification of the same periods at Fort Huachuca, the digital-photograph analysis often arrived at these temporal affiliations through the erroneous identification of types. In many cases in which the hands-on laboratory and digital-photograph analyses agreed as to the temporal periods associated with ceramic types, the identified ceramic types were different but dated to the same periods.

Table 22. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Temporal Period

Temporal Period (A.D.)	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
200–1250	—	—	—	—	—	—	—	—
900–1150	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**1.00**	0.33
1125–1300	0.25	—	0.25	0.17	**1.00**	**1.25**	0.25	**0.83**
1150–1450	0.25	0.25	0.38	0.29	—	—	0.38	0.13
1200–1400	**1.00**	**1.00**	—	**0.67**	—	—	—	—
Not determined	**1.00**	**1.00**	—	**0.67**	**1.25**	**0.75**	—	**0.67**
Total	0.45	0.40	0.25	0.37	0.45	0.40	0.25	0.37

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 23. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Temporal Period

Temporal Period (A.D.)	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
650–750	**1.00**	**1.00**	—	**0.67**	—	—	**3.00**	**1.00**
700–1300	**0.64**	**0.64**	**0.55**	**0.61**	**3.00**	**2.73**	**1.00**	**2.24**
750–950	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**1.00**	0.33
950–1150	**1.00**	**1.00**	**0.67**	**0.89**	0.07	—	—	0.02
950–1300	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
1150–1300	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**2.00**	**0.67**
Not determined	0.07	0.05	0.03	0.05	0.03	0.03	0.04	0.04
Total	0.14	0.12	0.09	0.11	0.13	0.12	0.09	0.11

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Temporal Precision

In order to further compare the temporal information that could potentially be derived from in-field, digital-photograph, and hands-on laboratory analyses, an additional metric was calculated: weighted average period length (Table 24). The weighted average period length provides an estimate of the precision of temporal affiliations that resulted from the ceramic identifications. The smaller the number, the higher the precision (no matter how accurate the identification). The metric was calculated in two ways. The first method involved using all the ceramic artifacts to calculate the metric. The second method involved calculating the same metric using an estimate of vessel count (or minimum number of vessels) rather than the combined total of all sherds.

At Fort Bliss, the weighted average period length exhibited only minor differences when the analysis results were compared using all sherds; the resulting weighted average period length was between 322 and 360 years. The difference in weighted average period length between the hands-on laboratory analysis and other analyses was less than 10 percent when all sherds were used in making the calculation.

When estimated vessel counts were used to perform the analysis, however, the results were quite different and suggested a smaller weighted average period length for the laboratory analysis than for the in-field analyses. Use of vessel counts for the analysis resulted in a weighted average period length of between 488 and 644 years. The difference between the laboratory analyses was less than 1 percent, but the difference between the hands-on laboratory analysis and the in-field analysis was between 17 and 32 percent. Thus, it would appear that if vessel counts were used to interpret chronology at the Fort Bliss sites, the laboratory analyses would be more precise, corresponding to shorter periods for temporal components than would be indicated by the in-field analysis.

At Fort Huachuca, the weighted average period length was substantially shorter for laboratory analyses than for in-field analysis, both when all ceramic artifacts were used to make the calculation and when vessel counts were used to make the calculation. Use of all sherds for the analysis resulted in a weighted average period length of between 346 and 600 years. The difference between the hands-on laboratory analysis and the other analyses was between 18 and 73 percent when all sherds were used to make the calculation.

When estimated vessel counts were used to perform the analysis at Fort Huachuca, the results were similar in overall pattern but suggested a smaller weighted average period length for several analyses and a larger difference between the hands-on laboratory analysis and the other analyses. Use of vessel counts for the analysis resulted in a weighted average period length of between 244 and 600 years. The difference between the laboratory analyses was more than 40 percent; the difference between the hands-on laboratory analysis and the in-field analysis for weighted average period length was between 117 and 145 percent. For the Fort Huachuca ceramic data, the implication of the results is that regardless of whether all sherds or vessel counts are used, the precision of temporal periods is, on average, greater for laboratory analyses (with shorter temporal periods more often identified). This is particularly the case for the hands-on laboratory analysis. The differences in temporal precision of ceramic identification between analyses were most pronounced when vessel counts were used. Thus, we may conclude that laboratory analysis of ceramic artifacts by a trained specialist, particularly hands-on laboratory analysis, can result in substantially more-precise temporal data than can be arrived at based on in-field analysis of ceramic artifacts.

Cohen's Kappa and McNemar's Test for Bias

As discussed in the introduction to this chapter, Cohen's kappa and McNemar's test for bias were calculated for select vessel elements, ceramic wares, and ceramic types. These statistics help to further evaluate the level of agreement between analyses in artifact identification and to identify the vessel elements, ceramic wares, and ceramic types for which there is evidence for systematic bias in artifact identification.

Table 24. Weighted Average Period Length, According to Analysis Type, by Site and Estimation Method

Estimation Method, by Site	Weighted Average Period Length (Years)				
	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Fort Bliss					
All sherds	341.7	322.1	353.1	359.6	343.0
Vessel estimate	643.8	575.0	489.3	487.5	547.6
Fort Huachuca					
All sherds	584.4	600.0	346.7	410.7	500.7
Vessel estimate	531.3	600.0	244.4	350.0	450.0

The statistics were calculated for the in-field and digital-photograph analyses with the assumption that the hands-on laboratory analysis was most likely to have been accurate.

Cohen's Kappa and McNemar's Test for Bias at Fort Bliss

At Fort Bliss, Cohen's kappa and McNemar's test for bias were calculated for body sherds, Jornada-region ceramics, and Three Rivers Red-on-terracotta, because these were the most commonly identified vessel element, ware, and ceramic type (Table 25). Cohen's kappa indicates that the level of agreement was good for the in-field analysis performed by Field Technician 1 and very good for the other analyses (see Table 3). Cohen's kappa was good for Jornada-region ceramics and for Three Rivers Red-on-terracotta only for the digital-photograph analysis. For Jornada-region ceramics, the level of agreement was poor for the in-field analysis performed by Field Technician 1 and fair for the in-field analysis performed by Field Technician 2. For Three Rivers Red-on-terracotta, the level of agreement was similar between in-field analyses and was either fair or moderate. McNemar's test for bias indicated that systematic bias was not evident for any of the compared types, despite the low level of agreement in a number of cases. Overall, an acceptable level of agreement was achieved for body sherds for all analyses and for Jornada-region ceramics and Three Rivers Red-on-terracotta only for the digital-photograph analysis.

Cohen's Kappa and McNemar's Test for Bias at Fort Huachuca

At Fort Huachuca, Cohen's kappa and McNemar's test for bias were calculated for body sherds, Hohokam Buff Ware, plain ware, indeterminate red-on-buff ware, Type II (smooth, sand), and Type III (micaceous schist) (Table 26). As with Fort Bliss, Cohen's kappa indicates a good or very good level of agreement for all analyses for body sherds. For many other types, however, the level of agreement is poor or fair, and evidence for systematic bias is common. For these types, an acceptable level of agreement was reached only for plain ware for the in-field analysis performed by Field Technician 1. All other analyses had unacceptable levels of agreement for all types. Moreover, evidence for systematic bias was common, particularly for in-field analyses. For digital-photograph analysis, systematic bias was evident for fewer types: Type II (smooth, sand) and Type III (micaceous schist). Systematic bias occurred for these types in the digital-photograph analysis, because the digital-photograph analysis tended to identify Type III (micaceous schist) sherds as Type II (smooth, sand) sherds and never identified an artifact as Type III (micaceous schist).

Table 25. Cohen's Kappa at Fort Bliss, According to Analysis Type, per Ceramic Type

Artifact Identification by Hands-on Laboratory Specialist	Observational Category	Cohen's Kappa		
		Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist
Body sherd	vessel element	0.63	0.89	0.89
Jornada-region ceramics	ceramic ware	0.08	0.27	0.63
Three Rivers Red-on-terracotta	ceramic type	0.39	0.47	0.69

Note: McNemar's test for bias indicated no systematic bias for these data.

Table 26. Cohen's Kappa at Fort Huachuca, According to Analysis Type, per Ceramic Type

Artifact Identification by Hands-on Laboratory Specialist	Observational Category	Cohen's Kappa		
		Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist
Body sherd	vessel element	0.85	0.79	0.77
Hohokam Buff Ware	ceramic ware	**0.17**	**0.18**	0.35
Plain ware	ceramic ware	0.75	**0.37**	0.40
Indeterminate red-on-buff ware	ceramic type	**0.15**	**0.17**	0.36
Type II (smooth, sand)	ceramic type	**0.46**	0.18	**0.13**
Type III (micaceous schist)	ceramic type	0.49	**0.32**	**—**

Note: Boldface values indicate systematic bias, based on the results of a McNemar's test for bias.

Summary and Discussion

The above discussions show that substantial variation in ceramic-artifact identification is likely to occur, depending on whether an analysis is performed in the field by a field technician, in the laboratory by a trained specialist using only digital photographs, or in the laboratory by a trained specialist who has direct access to the physical artifacts. Clearly, ceramic-artifact attributes were frequently misidentified by both in-field and digital-photograph analyses, sometimes at alarmingly high rates. Moreover, the levels of agreement between analyses generally declined as the number of potential attribute values increased. The highest levels of agreement were calculated for vessel element, of which only a few attributes are possible (e.g., rim sherd, body sherd, and indeterminate vessel element). Levels of agreement between analyses were lower for ceramic ware and still lower for ceramic type and temporal affiliation. This suggests that as the number of potential attribute values increases, variation among analyses in artifact identification also increases.

Interestingly, whether digital-photograph analysis performed better or worse than or similarly to in-field analysis varied among attributes as well as between sites. One might have expected digital-photograph analysis to perform better than in-field analysis, particularly because a trained specialist made the artifact identifications. It appears, however, that the digital photographs did not always adequately convey the information that was needed to correctly identify a ceramic artifact. Variation in artifact identification was also relatively common between in-field analyses, and a large divergence between field identifications was particularly common at Fort Huachuca. Variation among in-field analyses likely stemmed from differing understandings of artifact-type definitions as well as differing levels of experience among the field technicians. Overall, variation among analyses in artifact identification was greatest at Fort Huachuca—a result that appears to have stemmed from a larger and more diverse collection of ceramic artifacts that corresponded to a wider array of technological traditions and temporal affiliations and thus introduced a greater opportunity for errors to be made.

In many cases, difference among analyses in artifact identification would likely lead to substantial differences in site interpretation. It follows that at least some of these discrepancies also could have contributed to differences in how each site would be managed or evaluated. Some discrepancies among the analyses in artifact identification would have been of more consequence than others, such as the case in which a ceramic type at Fort Huachuca was identified by in-field analysis as an indeterminate red-on-buff sherd dating to A.D. 700–1300, although the ceramic was actually a Dragoon Red-on-brown sherd dating to A.D. 950–1150, as determined by hands-on laboratory analysis. In such a case, the in-field identification would suggest that the artifact is potentially affiliated with the Gila-Salt Basin Hohokam and dates to a broad period that encompasses the Pioneer, Colonial, Sedentary, and Early Classic periods of the Hohokam cultural sequence. The hands-on laboratory identification, in contrast, would indicate that the artifact is affiliated with the more locally based San Pedro River valley tradition and can be dated to a briefer period corresponding only to the Sedentary period of the Hohokam cultural sequence. In comparison to the in-field analysis, the hands-on laboratory analysis is arguably more accurate in terms of ceramic tradition, technology, and cultural affiliation as well as much more precise in terms of temporal affiliation. Obviously, these kinds of information are far more valuable to understanding the importance of a site and its relationship to other sites, in terms of site use, chronology, settlement pattern, and cultural affiliation.

By the same token, as substantial as some of the differences in artifact identification could be, the differences at Fort Bliss were not quite as disparate among analyses—a result that may correspond to less ceramic diversity at Fort Bliss than at Fort Huachuca as well as differences in the experience of field crews with recognizing ceramic-artifact types. It seems that it is possible in some investigative contexts that in-field analysis can arrive at ceramic-analysis results that are reasonably congruent with those from a hands-on laboratory analysis (at least in terms of their overall distribution), to the extent that a fairly accurate and precise interpretation of some aspects of a site's ceramic assemblage can be achieved through in-field analysis. However, this is not a guarantee. Clearly, broadly disparate results can also be obtained from different analytical settings and levels of experience.

As a consequence, a substantial level of uncertainty can be associated with interpreting the results of in-field analysis of ceramic artifacts. Although sometimes quite good, in-field analysis is likely to be less reliable than hands-on laboratory analysis. By the same token, relying on digital photographs alone for artifact analysis, even when analysis is performed by a trained specialist, can result in the erroneous identification of some ceramic wares and types, including those that are not likely to be present at a site. For instance, multiple red ware sherds were identified in the collection from the Fort Huachuca site by the digital-photograph analysis alone, but none was likely present in the collection. Similarly, a Galisteo Black-on-white ceramic artifact was identified in the collection from the Fort Bliss site by the digital-photograph analysis alone, but the sherd is likely to be Chupadero Black-on-white.

Tables 27 and 28 summarize some of the ceramic-analysis results that could have been used to interpret the Fort Bliss and Fort Huachuca sites investigated during this project. For both sites, the greatest number of vessels, highest ceramic richness, and greatest number of temporal periods could be interpreted based on the laboratory analyses. The greatest number of rim sherds was identified by the hands-on laboratory analysis. At both sites, the number of ceramic wares was likely overestimated by the digital-photograph analysis and underestimated by the in-field analysis.

At Fort Bliss, the minimum number of vessels and the number of ceramic types were likely overestimated by the digital-photograph analysis and underestimated by the in-field analysis. The number of rim sherds and the number of temporal periods were likely underestimated by the in-field and digital-photograph analyses. At Fort Huachuca, the numbers of rim sherds, ceramic vessels, ceramic types, and temporal periods were likely underestimated by the in-field and digital-photograph analyses, particularly the in-field analysis. At both sites, the precision of the inferred temporal periods was highest for laboratory analyses and considerably lower for in-field analysis, indicating that ceramic identifications made in the field would tend to identify sherds as belonging to types that would be placed in broader periods than those associated with types identified in the laboratory.

Table 27. Summary of Results of Ceramic Analysis at Fort Bliss, per Analysis Type

Observer	Rim-Sherd Count	Minimum Number of Vessels	Ceramic-Type Richness	Ceramic-Ware Richness	Temporal-Period Count	Weighted Average Temporal Period Length (Years), Based on Vessel Count	Ceramic Wares
Field Technician 1	4	8	7	3	3	575	Indeterminate painted ware; Jornada-region ceramics; plain ware.
Field Technician 2	6	8	6	2	3	529	Jornada-region ceramics.
Hands-on laboratory specialist	7	12	9	4	5	436	Chihuahua tradition; indeterminate painted ware; Jornada-region ceramics; plain ware.
Digital-photograph laboratory specialist	6	13	11	6	4	428	Chihuahua tradition; indeterminate painted ware; Jornada-region ceramics; **Mimbres Mogollon ceramics**; **Northern Rio Grande ware**; plain ware.

Note: Text in gray font indicates a ware that was identified by one analysis and was not identified by the hands-on laboratory analysis.

Table 28. Summary of Results of Ceramic Analysis at Fort Huachuca, per Analysis Type

Observer	Rim-Sherd Count	Minimum Number of Vessels	Ceramic-Type Richness	Ceramic-Ware Richness	Temporal-Period Count	Weighted Average Temporal Period Length (Years), Based on Vessel Count	Ceramic Wares
Field Technician 1	20	23	10	6	3	531	Hohokam Buff Ware; indeterminate painted ware; plain ware; San Pedro River valley tradition; **Trincheras**; Tucson Basin ceramics.
Field Technician 2	21	23	7	3	1	600	Hohokam Buff Ware; indeterminate painted ware; plain ware.
Hands-on laboratory specialist	24	35	20	8	6	309	Hohokam Buff Ware; indeterminate painted ware; plain ware; red ware; San Pedro River valley tradition; **San Simon series**; Tucson Basin ceramics; **Upper/Middle Santa Cruz River valley tradition**.
Digital-photograph laboratory specialist	22	32	18	9	5	345	**Brown or buff ware; brown or red ware; buff ware;** Hohokam Buff Ware; indeterminate painted ware; plain ware; red ware; San Pedro River valley tradition; Tucson Basin ceramics.

Note: Text in gray font indicates a ware that was identified by one analysis and was not identified by the hands-on laboratory analysis; text in boldface indicates a ware that was identified only by the hands-on laboratory analysis.

Finally, at both sites, the ceramic wares and types identified differed among analyses. At Fort Bliss, ceramic artifacts corresponding to Mimbres Mogollon ceramics and Northern Rio Grande wares were identified by the digital-photograph analysis but were not identified by any other analysis and were not likely present in the collection. At the Fort Huachuca site, a Trincheras ceramic artifact was identified by one of the in-field analyses, and ceramic types of several generic wares were identified by the digital-photograph analysis. Again, artifacts of these wares were likely not present in the collection. At the same time, the hands-on laboratory analysis identified ceramic artifacts of two relatively rare wares at Fort Huachuca—San Simon series and Upper/Middle Santa Cruz River valley tradition—that were not identified by any other analyses and would have been particularly important to an interpretation of the site's use during prehistory.

All together, these results suggest that the greatest level of accuracy and precision in ceramic-artifact analysis is achieved by hands-on analysis. Although some interpretations of a site resulting from in-field analysis or digital-photograph analysis will be in agreement with those derived from hands-on analysis, many diverge from interpretations that could be derived from hands-on analysis. We might therefore conclude that neither in-field analysis nor digital-photograph analysis are likely to be as reliable as hands-on laboratory analysis in identifying ceramic artifacts or deriving inferences based on those artifacts to interpret a site. Thus, evaluations or management decisions based on the results of in-field or digital-photograph analyses are more likely to be in error than those based on the results of hands-on laboratory analysis by a trained specialist.

CHAPTER 4

Lithic Analysis

In this chapter, the results of the four different analyses of lithic artifacts performed at each of the two project sites are discussed. For the Fort Bliss site, 105 lithic artifacts were analyzed. Most of the artifacts in the collection were flaked stone tools or cores or ground/battered stone tools. Though not rare at the site overall, flakes and debris were rare in the collection analyzed for this project. It appears that because absolute limits were placed on the numbers of artifacts subjected to analysis, investigators at Fort Bliss favored identification of tools and cores over identification of flakes. The number of lithic artifacts analyzed for the Fort Huachuca site was somewhat larger, totaling 150 artifacts; the majority of these artifacts were flakes. The absolute and relative frequencies of flaked stone tools and cores and ground/battered stone tools from the Fort Huachuca site were lower than those from the Fort Bliss site.

In this chapter, we have focused on making comparisons among analyses for a series of three attributes of lithic artifacts: production method (i.e., flaked stone, expedient use, ground/battered stone, or noncultural), material type, and technological type (e.g., biface, metate, or core). For each of these attributes, several metrics (as introduced in Chapter 2) were calculated: agreement index, false-negative and false-positive rates, Cohen's kappa, and McNemar's test for bias. Agreement indexes were calculated for all pairs of analyses in order to provide an overall understanding of which analyses were more or less in agreement. The other metrics were calculated by assuming that hands-on laboratory analysis provided the most-accurate results, and it functioned as the "gold standard" in artifact identification for this project.

In addition to the above metrics, the distributions of attribute values were compared among analyses in order to evaluate the degree to which site interpretations would converge or diverge if site interpretation were to be based on the results of the different individual analyses. Whereas the above metrics are based on per-artifact comparisons (either between pairs of analyses or between the hands-on analysis and the other analyses), the distributional comparisons investigated the potential implications that the different analyses could have in terms of the kinds of activities performed at a site and their relative intensities, regardless of whether individual artifact analyses were precisely the same or not. In other words, one analysis could identify six artifacts as cores made on rhyolite and three artifacts as flakes made on quartzite, whereas another analysis may identify a different set of six artifacts as cores made on rhyolite and a different set of three artifacts as flakes made on quartzite. Although, in this hypothetical scenario, none of the artifact identifications would agree with each other, the distribution of the results would be the same and could contribute in an identical manner to site interpretation, regardless of which specific artifacts were cores made on rhyolite or flakes made on quartzite.

Below, each of the lithic attributes focused on for analysis (production method, material type, and technological type) are discussed in turn. This is followed by a discussion of the results from calculations of Cohen's kappa and McNemar's test for bias for select artifact attributes and then by a summary of the results.

Production Method

In total, four production methods were recorded for artifacts made on lithic materials: expedient use, ground/battered stone, flaked stone, and noncultural. An expedient-use artifact is an artifact that has been culturally modified incidentally, as a result of use but not as a result of deliberate shaping or design. For this project, expedient-use artifacts are differentiated from flaked stone and ground stone artifacts that were intentionally shaped to fulfill a particular tool function. Generally, one would expect that the production method of lithic artifacts, as defined above, would be consistently identified, given that the categories correspond to artifacts of substantially divergent functions and manufacturing technologies. One might also expect there to be some disparity in terms of whether or not an artifact is identified as expedient use, because the assessment may, at times, amount to a judgment call regarding the level of effort put into preparing and shaping an artifact for use. However, expedient-use artifacts were either hammerstones or hand stones that had been used opportunistically for either percussion or grinding activities or both but that are sometimes misclassified as specific kinds of manos. The "noncultural" production method refers to collected artifacts that have been identified by a field technician as having been culturally modified but later identified as noncultural in origin by a lithic specialist during laboratory analysis.

Variation among analyses in production-method identification is discussed below for each of the two project sites, followed by a discussion of the agreement-index and misidentification-rate results for both sites.

Production Method at Fort Bliss

Two lithic artifacts from the Fort Bliss site were identified during the hands-on lithic analysis as noncultural (Table 29). The in-field and digital-photograph analyses, in contrast, considered the two artifacts to be cultural in origin. One of the artifacts was identified during the in-field or digital-photograph analysis as a core made on either chert or limestone. The other artifact was identified as either a tabular knife made on limestone or a retouched piece made on either chert or limestone.

Production-method identifications for the Fort Bliss site were frequently in close agreement among analyses, but they were never in perfect agreement. In addition to the two noncultural artifacts, identification of production methods for four artifacts varied among the analyses. One artifact (PD 12148) was identified as a flaked stone core by both in-field analyses but was identified as an expedient-use hammerstone by the digital-photograph analysis and as an expedient-use hand stone by the hands-on laboratory analysis. Another artifact (PD 12162) was identified by one in-field analysis as an expedient-use hammerstone but was identified by the other analyses as a ground/battered stone one-handed mano or undetermined mano. A third artifact (PD 12338) was identified as an expedient-use chert hammerstone by one of the in-field analyses but as a flaked stone limestone hammerstone by the other analyses. A fourth artifact (PD 12147) was identified as an expedient-use hand stone, an expedient-use hammerstone, or a flaked stone core. Although these variations in identification are understandable and probably would not have major consequences on interpretation of this site, variation in these identifications among observers could potentially have some effect on a site's interpretation if, for instance, such artifacts were the only instances of a particular production method at the site (i.e., flaked stone, ground/battered stone, or expedient use) and thus were used to identify the presence or absence of an activity involving such artifacts.

Production Method at Fort Huachuca

Five lithic artifacts from the Fort Huachuca site were identified during the hands-on lithic analysis as noncultural; one was also identified during the digital-photograph analysis as noncultural (Table 30). Among analyses that considered these noncultural specimens to be cultural in origin, classifications varied. One specimen was classified as a core flake, a tested cobble, or a retouched piece; a second was identified as either a core or a core flake; a third was identified as angular debris or noncultural; a fourth was identified as either a core flake or angular debris; and the fifth was identified as angular debris, a core, or FCR.

Classifications of 10 lithic artifacts from Fort Huachuca varied among the analyses (Table 31). These tended to be artifacts classified variously by analysts as flaked stone angular debris, FCR, cores, core flakes, or ground/battered stone objects of undetermined type. In one case, an artifact identified by the hands-on laboratory analysis as a flaked stone uniface was identified by the in-field analysis as a flaked

Table 29. Artifact Counts at Fort Bliss, According to Analysis Type, per Production Method

Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Expedient use	5	6	6	6	23
Flaked stone	78	78	75	77	308
Ground/battered stone	22	21	22	22	87
Noncultural	—	—	2	—	2
Total	105	105	105	105	420

Table 30. Artifact Counts at Fort Huachuca, According to Analysis Type, per Production Method

Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Expedient use	1	1	—	1	3
Flaked stone	143	146	139	142	570
Ground/battered stone	6	3	6	6	21
Noncultural	—	—	5	1	6
Total	150	150	150	150	600

Table 31. Cases of Production-Method Disagreement at Fort Huachuca, According to Analysis Type

Provenience Designation No.	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist
9	flaked stone quartzite fire-cracked rock	flaked stone quartzite angular debris	noncultural quartzite specimen	flaked stone quartzite core
51	flaked stone sandstone core flake	flaked stone andesite angular debris	noncultural quartzite specimen	flaked stone quartzite angular debris
121	flaked stone chert angular debris	flaked stone chalcedony angular debris	noncultural chert specimen	noncultural chert specimen
152	ground/battered stone quartzite undetermined ground stone	flaked stone basalt angular debris	ground/battered stone quartzite undetermined ground stone	ground/battered stone quartzite undetermined ground stone
247	expedient-use quartzite hammerstone	expedient-use quartzite hammerstone	ground/battered stone quartzite undetermined ground stone	ground/battered stone quartzite one-handed mano
263	ground/battered stone quartzite undetermined ground stone	flaked stone basalt angular debris	flaked stone quartzite core flake	flaked stone quartzite core flake
399	flaked stone rhyolite core	flaked stone other core	flaked stone rhyolite cobble uniface	expedient-use rhyolite hammerstone
429	flaked stone rhyolite core flake	flaked stone other core	noncultural rhyolite specimen	flaked stone rhyolite core flake
449	ground/battered stone quartzite undetermined ground stone	flaked stone quartzite core	ground/battered stone quartzite undetermined metate	ground/battered stone granite undetermined ground stone
479	flaked stone rhyolite retouched piece	flaked stone other core flake	noncultural rhyolite specimen	flaked stone rhyolite tested cobble

stone core and as an expedient-use hammerstone by the digital-photograph analysis. In another case, an artifact identified by the hands-on laboratory analysis as a ground/battered stone metate of undetermined type was identified by two analyses as a ground/battered artifact of undetermined type, and one of the in-field analyses identified the artifact as a flaked stone core.

Production-Method Agreement and Error Rates at Both Sites

Production-method identifications were generally consistent at both project sites, as one might expect. For Fort Bliss, the agreement index for production method was calculated as 0.97, on average (Table 32). For Fort Huachuca, the agreement index for production method was slightly lower and more variable, calculated as 0.96, on average (Table 33). In other words, between analyses, there was agreement as to the class of an artifact in more than 95 percent of cases.

Similarly, the false-positive and false-negative rates for production method were low, overall, for both sites: 0.05 or less for all analyses (Tables 34 and 35). For both sites, false-positive and false-negative rates were slightly higher overall for the in-field analysis than for the digital-photograph analysis. For Fort Bliss, artifacts identified by the hands-on laboratory analysis as expedient-use artifacts were occasionally misidentified by the in-field analysis as having been made through a different production method. In a couple of cases, a flaked stone or ground/battered stone artifact was identified by in-field analysis incorrectly as an expedient-use artifact. In contrast, the digital-photograph analysis always correctly identified whether an artifact was expedient use or not. At Fort Huachuca, noncultural artifacts were always misclassified by the in-field analysis; the digital-photograph analysis identified one artifact as noncultural that was also identified as noncultural by the hands-on laboratory analysis. For the Fort Huachuca site, artifacts identified as ground/battered stone by the hands-on laboratory analysis were misidentified by the in-field analysis in a number of cases but were never misidentified by the digital-photograph analysis. Flaked stone artifacts were rarely misidentified as having been made through a different production method. Artifacts identified as noncultural were misclassified as cultural in most cases at both sites.

Overall, misidentification rates for production method suggest that digital-photograph analysis accurately identified the production method of lithic artifacts nearly all of the time but that noncultural specimens were difficult to identify using only digital photographs. In-field analysis was also generally successful at correctly identifying the production method of an artifact nearly all of the time, but potentially, noncultural items were misidentified as cultural and ground/battered stone artifacts and expedient-use artifacts were misidentified in perhaps one-quarter or more of cases.

Material Type

The identification of lithic artifacts according to material type can be important for resolving a variety of issues related to site interpretation, including interpreting where site inhabitants may have traveled before arriving at a site, evidence for trade and exchange, material preference, territoriality, and sometimes even artifact function. For instance, vesicular basalt was often used in place of other materials to make trough metates for grinding maize, presumably because of the material's superior performance characteristics in that function (Adams 2002:Table 3.1). The identification of ground stone artifacts made on vesicular basalt, particularly trough metates and manos, can thus sometimes be important in inferring whether maize was likely to have been processed at a site, an important consideration for assessing site function as well as the contribution of cultigens to subsistence economies.

Table 32. Agreement Index for Lithic-Artifact Analysis at Fort Bliss, According to Production Method, Material Type, and Technological Type

Observer 1	Observer 2	Production-Method Agreement	Material-Type Agreement	Technological-Type Agreement	Overall Agreement
Hands-on laboratory specialist	Field Technician 1	0.98	0.72	0.54	0.75
Hands-on laboratory specialist	Field Technician 2	0.96	0.80	0.52	0.76
Hands-on laboratory specialist	Digital-photograph laboratory specialist	0.98	0.59	0.50	0.69
Field Technician 1	Field Technician 2	0.97	0.66	0.48	0.70
Field Technician 1	Digital-photograph laboratory specialist	0.97	0.68	0.48	0.71
Field Technician 2	Digital-photograph laboratory specialist	0.98	0.59	0.39	0.65
Average		0.97	0.67	0.48	0.71
CV (%)		0.8	12.0	10.7	5.8

Key: CV = coefficient of variation.

Table 33. Agreement Index for Lithic-Artifact Analysis at Fort Huachuca, According to Production Method, Material Type, and Technological Type

Observer 1	Observer 2	Production-Method Agreement	Material-Type Agreement	Technological-Type Agreement	Overall Agreement
Hands-on laboratory specialist	Field Technician 1	0.95	0.83	0.67	0.82
Hands-on laboratory specialist	Field Technician 2	0.95	0.09	0.33	0.46
Hands-on laboratory specialist	Digital-photograph laboratory specialist	0.97	0.79	0.65	0.80
Field Technician 1	Field Technician 2	0.98	0.10	0.37	0.48
Field Technician 1	Digital-photograph laboratory specialist	0.97	0.72	0.59	0.76
Field Technician 2	Digital-photograph laboratory specialist	0.97	0.09	0.31	0.46
Average		0.96	0.44	0.49	0.63
CV (%)		1.3	85.9	34.9	28.8

Key: CV = coefficient of variation.

Table 34. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Production Method

Production Method	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Expedient use	0.33	0.17	—	0.17	0.17	0.17	—	0.11
Flaked stone	0.01	—	—	—	0.05	0.04	0.03	0.04
Ground/battered stone	—	0.05	—	0.02	—	—	—	—
Noncultural	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Total	0.05	0.04	0.02	0.03	0.05	0.04	0.02	0.03

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 35. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Production Method

Production Method	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Flaked stone	0.01	—	0.01	—	0.04	0.05	0.03	0.04
Ground/battered stone	0.17	0.50	—	0.22	0.17	—	—	0.06
Noncultural	**1.00**	**1.00**	**0.80**	**0.93**	—	—	—	—
Total	0.05	0.05	0.03	0.04	0.04	0.05	0.03	0.04

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Material Type at Fort Bliss

At Fort Bliss, nine different material types were identified (Table 36). The material type for three artifacts was not determined by the hands-on laboratory analysis (i.e., was recorded as "undetermined"). The most commonly identified material types in the Fort Bliss collection were chert, limestone, and sandstone. Chalcedony was also identified in one or two instances by each analyst. Basalt, granite, obsidian, quartzite, and rhyolite were identified by three or fewer of the four analyses. Three of these material types (basalt, granite, and quartzite) were not identified by the hands-on laboratory analysis, suggesting that they may not have been present in the collection.

Interestingly, each analysis of lithic artifacts at Fort Bliss resulted in the identification of six or seven material types, but the identified material types varied per analysis. Thus, in-field analysis could potentially identify the correct number of material types in a collection of lithic artifacts, even when the material types for some artifacts are misidentified. In other words, field technicians can be effective in discerning

Table 36. Artifact Counts at Fort Bliss, According to Analysis Type, per Material Type

Material Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Basalt	—	—	—	1	1
Chalcedony	1	2	1	2	6
Chert	55	69	68	40	232
Granite	1	—	—	1	2
Limestone	29	20	14	53	116
Obsidian	—	—	1	—	1
Quartzite	1	3	—	—	4
Rhyolite	6	1	1	—	8
Sandstone	12	10	17	8	47
Not determined	—	—	3	—	3
Total	105	105	105	105	420
Richness	7	6	6	6	9

the presence of distinctive material types but may, at times, not be experienced enough to correctly identify what specific material types are present. An interesting result of the identification of lithic-material types at Fort Bliss was that, in comparison to other analyses, the digital-photograph analysis appears to have underemphasized the identification of chert and overemphasized the identification of limestone. The results for the digital-photograph analysis are highly anomalous in identifying more artifacts made on limestone and fewer artifacts made on chert than expected. When the results of the digital-photograph analysis for these two material types are compared to the results from the hands-on laboratory analysis using a chi-square test, the difference between the two sets of results is highly significant ($p < .0001$; $\chi^2 = 27.720$; $df = 1$) (see also the section on Cohen's Kappa and McNemar's test for bias at Fort Bliss, below). Because chert and limestone are often found in the same lithological deposits, it may be the case that it was difficult to differentiate chert materials from limestone materials using the digital photographs alone. Sandstone materials may also have been difficult to identify using the digital photographs, because artifacts identified as made on sandstone were least often identified as such by the digital-photograph analysis.

Another important result of the Fort Bliss lithic analysis is that only the hands-on laboratory analysis identified an artifact as made on obsidian. Because the provenance of artifacts made on obsidian can be determined using X-ray fluorescence, because obsidian artifacts can sometimes be dated using obsidian-hydration dating, and because, in some contexts, obsidian was preferred for the production of symbolically potent artifacts, the correct identification of artifacts as made on obsidian is especially important in the region (see Haury 1976; Hoffman 1997; Liritzis and Diakostamatiou 2002; Liritzis and Stevenson 2012; Riciputi et al. 2002; Shackley 2005). The artifact, which was identified by all analyses as a projectile point, was identified by the in-field and digital-photograph analyses as made on chert rather than obsidian.

At Fort Bliss, the agreement indexes for material type varied from 0.59 to 0.80, averaging 0.67 (see Table 32). The highest level of agreement was calculated for the comparison between the hands-on laboratory analysis and the in-field analysis performed by Field Technician 2. The lowest levels of agreement occurred between the digital-photograph analysis and the hands-on laboratory analysis and between the digital-photograph analysis and the in-field analysis performed by Field Technician 2.

If we assume that the hands-on laboratory analysis was the most accurate of the analyses in identifying material type, we can calculate false-negative and false-positive rates for material types identified by the hands-on laboratory analysis (Table 37). False-negative rates indicate that artifacts identified by the hands-on laboratory analysis as made on chalcedony, chert, or limestone were often identified correctly as

Table 37. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Material Type

Material Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Chalcedony	—	—	—	—	—	**1.00**	**1.00**	**0.67**
Chert	0.26	0.07	0.43	0.25	0.07	0.09	0.01	0.06
Limestone	0.14	0.29	—	0.14	**1.21**	**0.71**	**2.79**	**1.57**
Obsidian	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Rhyolite	**1.00**	—	**1.00**	**0.67**	**6.00**	—	—	**2.00**
Sandstone	0.29	0.47	**0.53**	0.43	—	0.06	—	0.02
Total	0.29	0.20	0.41	0.30	0.27	0.17	0.39	0.28

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

to material type. In other words, when the hands-on analysis identified an artifact as having been made on chalcedony, chert, or limestone, the in-field analysis and digital-photograph analysis typically also identified the artifact as having been made on the same material.

However, false-positive rates indicate that two of these material types, chalcedony and limestone, were frequently attributed to artifacts that were made on material types other than chalcedony or limestone. The same cannot be said for chert; artifacts identified by the hands-on laboratory analysis as made on a material other than chert were rarely identified incorrectly by in-field analysis or digital-photograph analysis as made on chert. In essence, a substantial number of artifacts made on chalcedony or limestone were usually identified correctly according to material type, but artifacts actually made on other materials tended to be identified as having been made on chalcedony or limestone. In a sense, one could say that analysts were often *accurate* in identifying chalcedony, chert, and limestone but were not *precise* in identifying chalcedony or limestone materials.

Artifacts identified by the hands-on analysis as made on rhyolite or obsidian, both rare material types in the collection, were identified as made on other material types all or most of the time. Unlike limestone or chalcedony, however, rhyolite and obsidian were, overall, rarely identified for artifacts made on materials other than rhyolite or obsidian, although Field Technician 1 did identify a number of artifacts that were not made on rhyolite as having been made on rhyolite (as indicated by a false-positive rate of 6). Artifacts identified by the hands-on analysis as made on sandstone were misidentified in approximately two of every five cases, overall. As with rhyolite and obsidian, artifacts made on materials other than sandstone were typically not identified by in-field or digital-photograph analysis as made on sandstone.

Interestingly, the highest false-negative and false-positive rates, overall, for material-type identification were calculated for the digital-photograph analysis. Digital-photograph analysis most often identified an artifact as made on limestone that was actually made on another material, such as chert. The digital-photograph analysis also tended more often to fail to correctly identify the material type of sandstone artifacts.

Material Type at Fort Huachuca

At Fort Huachuca, 12 material-type categories comprising 11 material types were identified by the different analyses (Table 38). One of the field technicians (Field Technician 2) identified material types very

Table 38. Artifact Counts at Fort Huachuca, According to Analysis Type, per Material Type

Material Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Andesite	3	2	—	—	5
Basalt	—	35	—	—	35
Basalt (vesicular)	—	1	—	1	2
Chalcedony	6	20	4	5	35
Chert	19	2	21	18	60
Diorite	—	—	1	—	1
Granite	—	—	1	1	2
Limestone	—	—	—	6	6
Quartzite	28	7	31	32	98
Rhyolite	91	5	91	84	271
Sandstone	1	—	—	2	3
Other	2	78	—	1	81
Not determined	—	—	1	—	1
Total	150	150	150	150	600
Richness	7	8	6	9	12

differently from the other analysts, which resulted in large disparities in material-type identification between the analysis performed by Field Technician 2 and all other analyses. Field Technician 2 identified the material type of an artifact as either basalt or "other" far more often than did any other analyst. Indeed, three-quarters of material-type identifications made by the analyst were one of these two material types. In contrast, no other analyst at Fort Huachuca identified an artifact as made on basalt (although in 1 case, another analysis identified an artifact as made on vesicular basalt), and in only a few cases did another analyst identify an artifact as made on an "other" material type. Field Technician 2 also identified a larger number of artifacts as made on chalcedony and fewer artifacts as made on chert, quartzite, or rhyolite than did the other analysts.

Moreover, materials identified by Field Technician 2 as basalt were identified by other analysts variously, most often as either rhyolite or quartzite. There was not a one-to-one correspondence between what Field Technician 2 called "basalt" and what the hands-on laboratory analyst called "rhyolite," for instance. Clearly, Field Technician 2 was not proficient in identifying material types discovered at the site, having identified more than half of the lithic artifacts as made on "other" materials and most of the rest as made on materials that were either absent or rare. Based on the hands-on-laboratory-analysis results, basalt appears to have been absent from the collection, and chalcedony was quite rare. Results of the other analyses suggest comparatively high levels of agreement for material-type identification.

At Fort Huachuca, the agreement indexes for material type (see Table 33) were much more variable among analyses than at Fort Bliss. This was largely because of the anomalous identifications made by Field Technician 2 for material type, as discussed above. The agreement indexes comparing material-type results between the in-field analysis performed by Field Technician 2 at Fort Huachuca and the material-type identifications made by other analyses were 0.1 or less for all interobserver comparisons. In contrast, the agreement indexes calculated for material-type comparisons among the other analyses at Fort Huachuca were much higher: greater than 0.7 in all cases. As a result of the wide disparity in agreement among analyses in terms of material type, the average agreement index for material type at Fort Huachuca was 0.44, indicating that, on average, fewer than half of the observations of material type were in agreement with each other. Excluding the in-field analysis performed by Field Technician 2, the average agreement index of comparisons among analyses was 0.78, indicating that material-type identifications were in agreement between observers for nearly four of five cases.

As with Fort Bliss, false-negative and false-positive rates can be calculated per material type, if we assume that the hands-on lithic analysis was most likely to have resulted in a correct identification (Table 39). Artifacts identified by the hands-on laboratory analysis as having been made on chalcedony were typically also identified by the other analyses as having been made on chalcedony. However, false-positive rates indicate that chalcedony was also identified as the material type for artifacts that were made on other materials. This was particularly the case for the analysis performed by Field Technician 2, who identified four times as many chalcedony artifacts than were actually present in the collection. The material type of an artifact made on diorite or granite was never correctly identified by the in-field or digital-photograph analysis, but the material type of an artifact not made on diorite or granite was rarely misidentified as made on either of these material types.

Artifacts made on the most-common material types—chert, quartzite, and rhyolite—were identified correctly most of the time for the digital-photograph analysis and the in-field analysis performed by Field Technician 1. Overall, false-negative and false-positive rates were low for these material types for the digital-photograph analysis and the in-field analysis performed by Field Technician 1. In contrast, false-negative rates were very high for Field Technician 2. The anomalous material-type identifications made by Field Technician 2 resulted in very high false-negative rates for artifacts made on chert, quartzite, or rhyolite. The high overall false-negative rate for Field Technician 2 is due to the frequent misclassification of artifacts as having been made on chalcedony, an “other” material type, or basalt, a material type not present in the collection.

If one were to rely only on the material-type identifications made by Field Technician 2 to interpret the lithic collection from the Fort Huachuca site, one would arrive at a very different interpretation of material use at the site than if one made the interpretation based on the results of the other analyses. Based on the in-field analysis performed by Field Technician 2, one might surmise that the lithic collection was dominated by artifacts made on basalt or chalcedony and that many other artifacts were made on one or more unusual material types that could not be identified (and were identified as “other”). Perhaps, such a result could prompt the need for petrographic analysis and the planning of lithic-source investigations. Based on the other analyses, however, an investigator would interpret the lithic collection as consisting mostly of artifacts made on chert, quartzite, or rhyolite. Variation in material-type identification among analyses could lead to differing interpretations of material preference as well as differing interpretations of potential source areas for lithic artifacts deposited at the site.

An interesting result of the material-type analysis was that Field Technician 1 performed best (and quite well) in accurately identifying the material types of lithic artifacts, suggesting that it is possible to obtain accurate identifications of material type from in-field analysis of lithic artifacts. However, the results also show that it is possible to obtain grossly inaccurate material-type identifications from in-field analysis. The digital-photograph analysis resulted in identifications that were fairly consistent with the hands-on laboratory identifications but also showed that some material types are difficult to differentiate using digital photographs alone.

Technological Type

Technological Type at Fort Bliss

In total, 23 different technological-type categories comprising 22 technological types were identified at Fort Bliss, including 12 flaked stone technological types (plus the category of “flaked stone other”), 2 expedient-use technological types, and 8 ground/battered stone technological types (Table 40). The richness of technological types was quite similar among analyses overall and for all production methods

Table 39. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Material Type

Material Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Chalcedony	0.25	—	—	0.08	**0.75**	**4.00**	0.25	**1.67**
Chert	0.24	**0.90**	0.19	0.44	0.14	—	0.05	0.06
Diorite	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Granite	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**1.00**	0.33
Quartzite	0.32	**0.77**	0.39	0.49	0.23	—	0.42	0.22
Rhyolite	0.08	**0.99**	0.14	0.40	0.08	0.04	0.07	0.06
Total	0.17	**0.91**	0.21	0.43	0.13	0.13	0.15	0.14

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 40. Artifact Counts at Fort Bliss, According to Analysis Type, per Technological Type

Technological Type, by Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Flaked stone					
Biface	6	15	11	5	37
Cobble biface	1	9	1	—	11
Cobble uniface	1	1	—	2	4
Core	23	23	26	20	92
Core flake	—	—	1	6	7
Hammerstone	—	1	1	1	3
Projectile point	3	1	3	2	9
Retouched piece	16	8	16	22	62
Scraper	10	12	2	—	24
Tabular knife	4	—	—	2	6
Tested cobble	2	2	—	1	5
Uniface	12	1	11	16	40
Other	—	5	3	—	8
Expedient use					
Hammerstone	5	6	3	6	20
Hand stone	—	—	3	—	3
Ground/battered stone					
Basin metate	—	—	—	3	3
Milling stone	—	1	—	5	6
One-handed mano	4	4	3	3	14
Slab metate	—	1	1	3	5

continued on next page

Technological Type, by Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Two-handed mano	4	2	1	1	8
Undetermined ground stone	—	—	1	—	1
Undetermined mano	7	5	10	6	28
Undetermined metate	7	8	6	1	22
Noncultural	—	—	2	—	2
Total	105	105	105	105	420
Richness, flaked stone	10	11	10	10	13
Richness, expedient use	1	1	2	1	2
Richness, ground/battered stone	4	6	6	7	8
Richness, all classes	15	18	18	18	23

but was lower overall for the analysis performed by Field Technician 1. Thus, based on the results of the in-field analysis performed by Field Technician 1, one might conclude that activities involving lithic technology were somewhat less diverse at the site than one would conclude based on the results of the other analyses of lithic artifacts. The number of flaked stone tools, flaked stone cores, manos, and metates were also broadly similar among analyses.

However, two of the analyses (the in-field analysis performed by Field Technician 2 and the digital-photograph analysis) identified fewer manos and more metates or milling stones than were likely present. The digital-photograph and in-field analyses also identified more flaked stone tools and fewer flaked stone cores than were likely present. Such differences in analysis results might not have major effects on the interpretation of lithic technology in terms of the general activities performed, but they could effect interpretations regarding the intensity of activities involving lithic tools as well as the degree of emphasis placed on different activities. Results of the in-field or digital-photograph analysis, for instance, could potentially be interpreted to indicate that there was a slightly greater emphasis on tool use or maintenance, as opposed to core reduction, than actually may have been the case (Table 41).

Despite some general similarities in the results of the analyses, there were frequent discrepancies between analyses in the identification of specific technological types, suggesting that there could be substantially different interpretations of the specific kinds of lithic technologies represented at the site. For instance, the number of bifaces and cobble bifaces identified at the site varied widely, which could lead to different interpretations about the importance of bifacial technology at the site. Although projectile points were rare, even the numbers of projectile points varied among analyses, with anywhere from one to three projectile points identified by the different analyses. Whether an artifact was identified as a scraper or a uniface, or more simply as a retouched piece, also varied widely among analyses. In addition, hammerstones were more commonly identified by the in-field and digital-photograph analyses than by the hands-on laboratory analysis, which identified some of the same artifacts as manos or hand stones. These results could potentially be used to infer a greater emphasis on flaked stone technology, as opposed to ground stone technology, than actually may have been the case.

Two of the analyses (the in-field analysis performed by Field Technician 1 and the digital-photograph analysis) identified the presence of tabular knives at the site, whereas the hands-on laboratory analysis suggested that no such artifacts were present in the collection. Because tabular knives are often interpreted as having been used aboriginally in the American Southwest to procure and process agave (Doolittle and Neely 2004), the identification of tabular knives in the collection could lead to the inference that activities focused on agave procurement and processing were performed by users of the site.

As for ground/battered stone artifacts, the digital-photograph analysis, in particular, identified basin metates, milling stones, and slab metates substantially more often than the other analyses, a result that could result in erroneous conclusions regarding the kinds of ground stone technologies employed at the

Table 41. Artifact Counts at Fort Bliss, According to Analysis Type, per General Technological Type

Artifact Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Mano	15	11	14	10	50
Metate	7	9	7	7	30
Hand stone	—	—	3	—	3
Milling stone	—	1	—	5	6
Hammerstone	5	7	4	7	23
Flake	—	—	1	6	7
Flaked stone core	23	23	26	20	92
Flaked stone tool	53	47	44	49	193
Other (including noncultural specimens)	2	7	6	1	16
Total	105	105	105	105	420

site as well as the kinds of materials processed using ground stone. Two-handed manos, which may have been used for more-intensive grinding activities and were often used aboriginally in trough metates to process maize (Adams 2002), were more often identified by the in-field analysis than by the laboratory analyses. One-handed manos were also identified slightly more often by the in-field analysis.

At Fort Bliss, technological-type agreement ranged from 0.39 to 0.54, averaging 0.48 (see Table 32). Agreement was highest between the hands-on laboratory analysis and the other analyses and was lowest between in-field analyses and between the digital-photograph and in-field analyses.

Overall, false-negative rates were higher than false-positive rates for technological type (Table 42). Technological types with the lowest false-negative rates included cores, hammerstones, one-handed manos, and two-handed manos; technological types with the lowest false-positive rates included cores, hand stones, projectile points, undetermined ground stone, undetermined manos, undetermined metates, other, and noncultural. As discussed above, artifacts of many types were frequently misclassified, a result that likely would have led to erroneous conclusions about the characteristics of lithic technology at the site. For instance, the digital-photograph analysis misclassified undetermined metates as more-definitive technological types, including basin metates, milling stones, and slab metates. The in-field analysis made this error far less frequently than the digital-photograph analysis.

When technological type was misclassified, the erroneous technological types provided by in-field or digital-photograph analysis were diverse. For instance, artifacts identified by the hands-on laboratory analysis as retouched pieces were identified variously by other analyses as unifaces, scrapers, core flakes, cores, tabular knives, and bifaces. Artifacts identified by the hands-on laboratory analysis as bifaces were identified by other analyses as unifaces, retouched pieces, and scrapers. Artifacts identified by the hands-on laboratory analysis as projectile points were identified by other analyses as bifaces. Together, these data suggest that misclassification of lithic artifacts according to type can be quite common. Some of these differences are understandable, given some overarching morphological similarities between technological types (i.e., bifaces and projectile points).

Technological Type at Fort Huachuca

In total, 19 different technological types were identified at Fort Huachuca, including 14 flaked stone technological types, 1 expedient-use technological type, and 4 ground/battered stone technological types (Table 43). The richness of technological types varied somewhat among analyses; greater numbers of types

Table 42. False-Negative and False-Positive Error Rates at Fort Bliss, According to Analysis Type, per Technological Type

Technological Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Biface	**0.64**	0.18	**0.64**	0.48	0.18	**0.55**	0.09	0.27
Cobble biface	**1.00**	**1.00**	**1.00**	**1.00**	**1.00**	**9.00**	—	**3.33**
Core	0.31	0.27	0.38	0.32	0.19	0.15	0.15	0.17
Core flake	**1.00**	**1.00**	—	**0.67**	—	—	**5.00**	**1.67**
Hammerstone	—	—	—	—	0.25	**0.75**	**0.75**	**0.58**
Hand stone	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
One-handed mano	0.33	0.33	0.33	0.33	**0.67**	**0.67**	0.33	**0.56**
Projectile point	0.33	**0.67**	0.33	0.44	0.33	—	—	0.11
Retouched piece	**0.56**	**0.63**	0.44	**0.54**	**0.56**	0.13	**0.81**	0.50
Scraper	0.50	**1.00**	**1.00**	**0.83**	**4.50**	**6.00**	—	**3.50**
Slab metate	**1.00**	**1.00**	**1.00**	**1.00**	—	**1.00**	**3.00**	**1.33**
Two-handed mano	—	—	—	—	**3.00**	**1.00**	—	**1.33**
Undetermined ground stone	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Undetermined mano	0.40	0.50	0.40	0.43	0.10	—	—	0.03
Undetermined metate	0.17	0.17	**0.83**	0.39	0.33	0.50	—	0.28
Uniface	0.45	**1.00**	0.45	**0.64**	**0.55**	0.09	**0.91**	**0.52**
Other	**1.00**	0.33	**1.00**	**0.78**	—	**1.00**	—	0.33
Noncultural	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Total	0.47	0.49	0.50	0.49	0.40	0.45	0.38	0.41

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

Table 43. Artifact Counts at Fort Huachuca, According to Analysis Type, per Technological Type

Technological Type, by Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Flaked stone					
Angular debris	7	32	—	13	52
Biface	—	—	1	—	1
Cobble biface	4	1	—	1	6
Cobble uniface	1	—	3	2	6
Core	10	12	12	7	41

Technological Type, by Production Method	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Core flake	90	42	112	89	333
Core tablet	—	—	2	—	2
Fire-cracked rock	1	—	—	—	1
Hammerstone	—	—	—	1	1
Projectile point	2	2	2	2	8
Retouched piece	14	—	4	18	36
Tested cobble	1	—	—	3	4
Undetermined flake	11	56	3	4	74
Uniface	2	1	—	2	5
Expedient use					
Hammerstone	1	1	—	1	3
Ground/battered stone					
One-handed mano	—	—	—	2	2
Undetermined ground stone	4	1	4	2	11
Undetermined mano	2	2	1	1	6
Undetermined metate	—	—	1	1	2
Noncultural	—	—	5	1	6
Total	150	150	150	150	600
Richness, flaked stone	11	7	8	11	14
Richness, expedient use	1	1	—	1	1
Richness, ground/battered stone	2	2	3	4	4
Richness, all classes	14	10	11	16	19

than were identified by the hands-on laboratory analysis were identified by the in-field analysis performed by Field Technician 1 and the digital-photograph analysis. The number of types identified during the in-field analysis by Field Technician 2 was one fewer than the number identified during the hands-on laboratory analysis. Potentially, an investigator might infer a somewhat different diversity of activities involving lithic technology as having occurred at the site, depending on which analysis was used to interpret the site.

Technological types identified during either the in-field or the digital-photograph analysis that were not identified during the hands-on laboratory analysis were angular debris, cobble bifaces, FCR, tested cobbles, unifaces, and one-handed manos. The greatest number of these artifacts was identified as angular debris. The same artifacts were typically identified by the hands-on laboratory analysis as core flakes or, less often, cores. In a few cases, particularly in the case of artifacts identified by Field Technician 2 as angular debris, the same artifacts were identified by the hands-on laboratory analysis as retouched pieces, ground stone artifacts of undetermined type, or noncultural specimens.

Technological types that were unique to the identifications made by the hands-on laboratory analysis were biface and core tablet (a flake removed across the worn platform surface of a core to create a new, rejuvenated platform surface). The artifact identified by the hands-on laboratory analysis as a biface was identified by the other analyses as a core. The two artifacts identified by the hands-on lithic analysis as core tablets were identified variously by the other analyses as cores, core flakes, or retouched pieces.

In many cases, the differences in identification are understandable, because they correspond to technological types of similar morphology that could overlap in function. Some of the differences could simply be chalked up to differences of opinion or subtle variations in the interpretation of technological-type definitions. In some ways, the differences in identification between analyses may not amount to great differences in terms of the lithic technologies implemented at the site. At the same time, the differences in identification would result in substantial differences in the numbers of tools, flakes, and angular debris at the site (Table 44). The number of flaked stone cores also differed between analyses, but not by a large amount.

Depending on the analysis used to interpret the site, an investigator might conclude that there was anywhere from 4 (based on the in-field analysis performed by Field Technician 2) to over 20 (based on the in-field analysis performed by Field Technician 1 or the digital-photograph analysis) lithic tools at the site. In contrast, based on the results of the hands-on laboratory analysis, an investigator might conclude that on the order of 10 lithic tools were present in the collection.

Similarly, the proportions of lithic tools, core flakes, and angular debris varied substantially among analyses. The in-field analysis performed by Field Technician 1 and the digital-photograph analysis resulted in the identification of the largest numbers of tools, a large number of core flakes, and a relatively small number of angular-debris artifacts. In contrast, the analysis performed by Field Technician 2 resulted in a relatively large number of angular-debris artifacts, fewer core flakes, and small numbers of tools. The hands-on laboratory analysis resulted in the identification of no angular-debris artifacts, the largest number of flakes, and moderate numbers of tools. The differences among analyses could potentially lead to differing interpretations regarding the nature of core reduction performed at the site and the degree of emphasis placed on primary core reduction vs. tool manufacture, maintenance, or discard.

Because of the small number of ground/battered stone artifacts identified in the collection by any analyst at Fort Huachuca and the relatively nondescript nature of the collected ground/battered stone artifacts, interpretation of ground stone technology at the site would probably not vary considerably depending on which analysis results were used to interpret the site. However, it is worth noting that only the laboratory analyses identified a ground/battered stone artifact as a metate, although the two laboratory analyses did not identify the same artifact as a metate. Artifacts identified as metates during either of the laboratory analyses were identified in the field as either ground stone artifacts of undetermined type or as cores.

At Fort Huachuca, the levels of agreement between different analyses in technological-type identifications were much more variable than at Fort Bliss (see Table 33). The agreement indexes for technological-type identifications ranged from 0.31 to 0.67, averaging 0.49. The lowest levels of agreement were calculated between the in-field analysis performed by Field Technician 2 and the other analyses. Only about one-third of the technological-type identifications made during the in-field analysis performed by Field Technician 2 were consistent with the technological-type identifications made by the other analysts. In contrast, the levels of agreement among the other analyses ranged from 0.59 to 0.67.

Calculation of false-negative and false-positive rates also showed that technological types were frequently misclassified at Fort Huachuca (Table 45). As would be expected, given the above discussion, the false-negative and false-positive rates were generally lower for the digital-photograph analysis and the in-field analysis performed by Field Technician 1 than for the in-field analysis performed by Field Technician 2. More than the other analysts, Field Technician 2 frequently identified a technological type as absent when it was actually present, and this occurred for a wide range of technological types. Field Technician 2 also identified a technological type as present when it was absent more often than other analysts, but to a lesser degree. This kind of error occurred most frequently for artifacts identified by Field Technician 2 as undetermined flakes.

In general, the lowest false-negative rates were calculated for core flakes, projectile points, undetermined flakes, and undetermined manos. Bifaces, cobble unifaces, core tablets, and undetermined metates were always misidentified (i.e., no analyst correctly identified such artifacts in the collection). Cores, retouched pieces, undetermined ground stone artifacts, and noncultural specimens were identified as absent in most cases in which they were actually present.

The most-substantial difference between the in-field and digital-photograph analyses, in terms of misidentifications, was for artifacts identified by the hands-on laboratory analysis as core flakes, a technological type that composed three-quarters of the artifact identifications made by the hands-on laboratory analysis. The in-field analysis performed by Field Technician 1 and the digital-photograph analysis each failed to identify core flakes correctly in less than 25 percent of cases. In contrast, more than two-thirds of

Table 44. Artifact Counts at Fort Huachuca, According to Analysis Type, per General Technological Type

Artifact Type	Field Technician 1	Field Technician 2	Hands-on Laboratory Specialist	Digital-Photograph Laboratory Specialist	Total
Angular debris	7	32	—	13	52
Mano	2	2	1	3	8
Metate	—	—	1	1	2
Hammerstone	1	1	—	2	4
Flake	101	98	115	93	407
Flaked stone core	11	12	14	10	47
Flaked stone tool	23	4	10	25	62
Other (including noncultural specimens)	5	1	9	3	18
Total	150	150	150	150	600

Table 45. False-Negative and False-Positive Error Rates at Fort Huachuca, According to Analysis Type, per Technological Type

Technological Type	False-Negative Rate				False-Positive Rate			
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist	Total
Biface	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Cobble uniface	**1.00**	**1.00**	**1.00**	**1.00**	0.33	—	**0.67**	0.33
Core	0.50	0.50	**0.67**	**0.56**	0.33	0.50	0.25	0.36
Core flake	0.24	**0.68**	0.24	0.39	0.04	0.05	0.04	0.04
Core tablet	**1.00**	**1.00**	**1.00**	**1.00**	—	—	—	—
Noncultural	**1.00**	**1.00**	**0.80**	**0.93**	—	—	—	—
Projectile point	—	—	—	—	—	—	—	—
Retouched piece	0.50	**1.00**	0.25	**0.58**	**3.00**	—	**3.75**	**2.25**
Undetermined flake	0.33	—	**0.67**	0.33	**3.00**	**17.67**	**1.00**	**7.22**
Undetermined ground stone	0.50	**0.75**	**0.75**	**0.67**	0.50	—	0.25	0.25
Undetermined mano	—	—	—	—	**1.00**	**1.00**	—	**0.67**
Undetermined metate	**1.00**	**1.00**	**1.00**	**1.00**	—	—	**1.00**	0.33
Total	0.33	**0.67**	0.35	0.45	0.23	0.44	0.19	0.29

Note: Cells shaded in gray indicate low rates, and cells without shading indicate either moderate rates (normal typeface) or high rates (boldface).

core flakes were misidentified by the in-field analysis performed by Field Technician 2, who most often identified these artifacts as angular debris.

Cohen's Kappa and McNemar's Test for Bias

As discussed in the introduction to this chapter, Cohen's kappa and McNemar's test for bias were calculated for select production methods, material types, and technological types. These statistics help to further evaluate the adequacy of the levels of agreement between analyses in artifact identification as well as to identify production methods, material types, or technological types for which there is evidence for systematic bias in artifact identification.

Cohen's Kappa and McNemar's Test for Bias at Fort Bliss

At Fort Bliss, Cohen's kappa and McNemar's test for bias were calculated for the production methods of flaked stone and ground/battered stone as well as for some of the more commonly identified material and technological types. Table 46 shows that the level of agreement for production method was very good (see Table 31) and was best for the production method of ground/battered stone. McNemar's test for bias indicated that systematic bias did not likely affect production-method identifications.

In comparison to the calculations for production method, Cohen's kappa was lower for commonly identified material types (chert, limestone, and sandstone), although identifications for these material types were generally moderate to very good. More than half of the comparisons between analyses indicated that an acceptable level of agreement occurred for a specific material type ($\kappa \geq 0.5$). However, there were often only moderate levels of agreement between the hands-on laboratory analysis and the other analyses for chert or limestone materials; unacceptable levels of agreement and evidence for systematic bias occurred for these material types for the digital-photograph analysis (for chert and limestone) and for the in-field analysis performed by Field Technician 2 (for limestone only).

The identification of an artifact as a specific kind of mano or metate was relatively rare at Fort Bliss. Therefore, Cohen's kappa and McNemar's test for bias had to be calculated using generalized classifications encompassing all manos or all metates (including both specific and undetermined types). Interestingly, although there appears to have been disagreement over the kind of mano or metate represented by an artifact (see above discussion on technological types at Fort Bliss), there seems to have been good to very good agreement regarding whether an artifact was, more generally, a mano or a metate; there was no evidence for systematic bias in making such identifications. Thus, although in-field or digital-photograph analysis may not produce very reliable results regarding whether a mano is a one-handed mano, a two-handed mano, or a mano of undetermined type, it can produce reliable results regarding whether a lithic artifact is, more simply, a mano or a metate and not some other kind of artifact.

For Fort Bliss, Cohen's kappa and McNemar's test for bias were calculated for the flaked stone technological types of core, biface, retouched piece, scraper, and uniface. The same statistics were also calculated for the more-general category of retouched tool, which, in this case, included bifaces, projectile points, retouched pieces, scrapers, tabular knives, and unifaces. For the in-field and digital-photograph analyses, Cohen's kappa indicated that there was a good level of agreement in identifying an artifact as a core, and there was no evidence for systematic bias. However, with the exception of artifacts identified as bifaces by the in-field analysis performed by Field Technician 2, the agreement between the hands-on laboratory analysis and the other analyses in the identification of an artifact as a biface, a retouched piece, a scraper, or a uniface was generally poor or fair. Cohen's kappa was below the threshold of 0.5 (an acceptable level of agreement) in artifact identification for all these comparisons but one, and McNemar's test for bias indicated the occurrence of systematic bias for artifacts identified as scrapers or unifaces. As discussed in the above section on technological types at Fort Bliss, different observers tended to identify

Table 46. Cohen's Kappa at Fort Bliss, According to Analysis Type, per Lithic-Artifact Type

Artifact Identification by Hands-on Laboratory Specialist, by Observational Category	Cohen's Kappa		
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist
Production method			
Flaked stone	0.88	0.93	0.95
Ground/battered stone	1.00	0.97	1.00
Material type			
Chert	0.55	0.76	**0.45**
Limestone	**0.46**	0.51	**0.26**
Sandstone	0.80	0.62	0.60
Technological type			
Ground/battered stone			
Mano[a]	0.96	0.86	0.81
Metate[a]	0.85	0.81	0.71
Flaked stone			
Core	0.65	0.71	0.61
Biface	0.43	0.65	0.46
Retouched piece	0.34	0.44	0.36
Scraper	**0.14**	**–0.03**	0.00
Uniface	0.46	**–0.02**	0.37
Retouched tool[a]	0.81	0.80	0.77

Note: Boldface values indicate systematic bias, based on the results of a McNemar's test for bias.
[a] Collapsed category including multiple identified artifact types.

scrapers or unifaces significantly more or less often than the hands-on laboratory analysis indicated that they were likely to have been present in the collection.

Interestingly, identifying more simply whether an artifact was a retouched tool or not (regardless of whether it was a biface, a projectile point, a retouched piece, a scraper, a tabular knife, or a uniface) was, in contrast, good to very good for the in-field and digital-photograph analyses. Thus, we may conclude that identification of an artifact as a retouched tool by in-field or digital-photograph analysis could be reliable, but determination of the specific kind of retouched tool represented by an artifact by in-field or digital-photograph analysis is likely to be unreliable.

Cohen's Kappa and McNemar's Test for Bias at Fort Huachuca

At Fort Huachuca, Cohen's kappa and McNemar's test for bias were calculated for the production methods of flaked stone as well as for some of the more commonly identified material and technological types. Table 47 shows that the level of agreement for the production method of flaked stone was moderate to good for the in-field and digital-photograph analyses. Identification of artifacts according to material type for the most-common material types of chert, rhyolite, and quartzite was moderate to very good for the digital-photograph analysis and the in-field analysis performed by Field Technician 1 but was poor to fair for the in-field analysis performed by Field Technician 2. There is also evidence for systematic bias in the identification of all three material types by Field Technician 2. Based on the results from the Fort Huachuca site, we might conclude that the identification of common material types by in-field or digital-photograph

Table 47. Cohen's Kappa at Fort Huachuca, According to Analysis Type, per Lithic Artifact Type

Artifact Identification by Hands-on Laboratory Specialist, by Observational Category	Cohen's Kappa		
	Field Technician 1	Field Technician 2	Digital-Photograph Laboratory Specialist
Production method			
Flaked stone	0.65	0.51	0.72
Material type			
Chert	0.77	**0.15**	0.85
Quartzite	0.64	**0.32**	0.50
Rhyolite	0.80	**–0.05**	0.74
Technological type (flaked stone)			
Core	0.51	0.46	0.38
Core flake	0.53	**0.10**	0.54
Undetermined flake	**0.26**	**0.07**	0.27
Retouched tool[a]	**0.36**	0.59	**0.37**

Note: Boldface values indicate systematic bias, based on the results of a McNemar's test for bias.
[a] Collapsed category including multiple identified artifact types.

analysis can be reliable but can also be highly unreliable, as was suggested in the above discussion of material-type identifications at Fort Huachuca. Field Technician 2 tended to identify the material type of an artifact as basalt, chalcedony, or "other" instead of chert, rhyolite, or quartzite, the material types more commonly identified by the other analyses of the same artifacts.

Because much of the collection from the Fort Huachuca site consisted of flakes and, to a lesser extent, cores, Cohen's kappa and McNemar's test for bias were calculated for the technological types of core, core flake, and undetermined flake. Unlike at Fort Bliss, ground/battered stone tools were found in small numbers overall; specific kinds of retouched flaked stone tools in the collection from Fort Huachuca were also rare. As with Fort Bliss, Cohen's kappa and McNemar's test for bias were also calculated for the more-general category of retouched tools, which, in this case, included bifaces, retouched pieces, projectile points, and unifaces.

The identification of cores at Fort Huachuca was fair to moderate; cores were identified at an acceptable level of agreement with the hands-on laboratory analysis only by the in-field analysis performed by Field Technician 1, and the level of agreement was only barely acceptable. Core flakes were identified at moderate and acceptable levels by the analysis performed by Field Technician 1 and by the digital-photograph analysis; there was a poor level of agreement between the hands-on analysis and the in-field analysis performed by Field Technician 2 for the identification of core flakes, as well as evidence for systematic bias. Field Technician 2, as discussed above in the section on technological types at Fort Huachuca, often identified core flakes as angular debris. There were poor to fair levels of agreement for the identification of undetermined flakes for all in-field and digital-photograph analyses, and there was evidence of systematic bias for two of the analyses.

Discouragingly, Cohen's kappa indicated that an acceptable level of agreement between the hands-on laboratory analysis and the other analyses was only achieved by one analysis for the general category of retouched tools. Ironically, the in-field analysis performed by Field Technician 2 achieved an acceptable level of agreement for the general category of retouched tools, although the levels of agreement for other identifications made by Field Technician 2 were often the lowest of all of the analyses. McNemar's test for bias also showed no evidence for systematic bias for this analysis. Despite this, the in-field analysis performed by Field Technician 2 was no more successful than the other analyses in correctly identifying artifacts that were identified as retouched tools by the hands-on laboratory analysis. In other words, the analysis had a relatively high false-negative rate as well as a relatively low false-positive rate for retouched tools.

The level of agreement was unacceptably low for the other analyses for the general category of retouched tools. McNemar's test for bias indicated that there was systematic bias in the identification of retouched tools for both the digital-photograph analysis and the in-field analysis performed by Field Technician 1. Both of these analyses tended to incorrectly identify as a retouched piece an artifact that was not identified by the hands-on laboratory analysis as having been retouched. In other words, false-positive rates for retouched tools at Fort Huachuca were high, a result that could lead to the interpretation of more-intensive tool use at the site than actually may have been the case.

As discussed above, the results from the Fort Bliss site indicated that the identification of specific kinds of retouched tools could be poor, but the general identification of an artifact as a retouched tool could be very good. The results from Fort Huachuca suggest, conversely, that it is possible for both in-field and digital-photograph analyses to identify substantially more retouched tools than were likely to be present at a site. It would seem, then, that whether the identification of an artifact as a retouched tool in the field or using a digital photograph could be considered reliable likely depends on a variety of factors, including the experience of field crew, the material type, and perhaps the tool type.

Summary and Conclusions

Together, the results of the analyses discussed above suggest that there can be a great amount of variability in the levels of agreement between analyses in the identification of lithic artifacts. As one might expect, the identification of a lithic artifact according to the general category of production method can be quite good. Although it can also be difficult for in-field or digital-photograph analysis to discern expedient-use artifacts, this was not necessarily much of a problem, because the production methods for the vast majority of artifacts were accurately identified as flaked stone or ground/battered stone.

The identification of the material type of a lithic artifact can be quite variable; some analyses achieved a relatively high level of agreement with the hands-on laboratory analysis, whereas others demonstrated low levels of agreement with the hands-on laboratory analysis. Systematic bias in the identification of materials types was evident for some material types and analyses. In the case of the in-field analysis performed by Field Technician 2 at Fort Huachuca, most lithic artifacts were incorrectly identified according to material type, and many were placed into the broad, ambiguous material-type category of "other." In contrast, the hands-on laboratory analysis showed that most artifacts in the collection from Fort Huachuca could be identified according to a few relatively common material types. Overall, material-type identifications tended to agree for less than two-thirds of cases at Fort Bliss and less than half of cases at Fort Huachuca. Some analyses performed well at identifying the material types of lithic artifacts, although rare material types, such as diorite or granite, tended to be misidentified. As discussed above, it is also possible for some analysts to misidentify common material types most of the time.

The lowest level of agreement was calculated for technological type. Technological-type identifications were in agreement with each other for approximately 50 percent of cases or fewer at both sites and were never greater than about two-thirds of cases, overall. In some ways, this was as expected, because there were more potential categories within which to place artifacts, and in some cases, typological definitions can be overlapping or fuzzy. In general, the more specific a technological-type classification, the less likely it is that agreement will be achieved between analyses. Analyses identified different kinds of manos or metates, for instance, but observers tended to agree more generally that an artifact was a mano, a metate, or another technological type. At Fort Bliss, different observers tended to agree as to what constituted a retouched tool but not as to the specific kind of retouched tool. At Fort Huachuca, different observers tended to disagree as to what constituted a retouched tool. There, different observers tended to read more or less into the design of an artifact, one seeing an artifact as a retouched tool and another seeing the same artifact as simply a flake or a core.

To some degree, some of these differences in identification are understandable, but they also indicate that substantial differences in site interpretation could result from an analysis, depending on who performs the analysis and in what context it is performed. Like the ceramic-analysis results, the lithic-analysis results also showed that in-field analysis of artifacts can be as good as or better than digital-photograph analysis and that digital-photograph analysis can suffer from problems that likely stem from an analyst's being unable to examine the physical artifact.

CHAPTER 5

Summary and Recommendations

Throughout the western United States, limited-collection and no-collection policies promulgated by federal agencies and SHPOs have led to the widespread use of in-field analysis for analyzing most or all artifacts used to characterize and interpret a site. This trend counters a century-long practice of collecting samples of surface artifacts to assist in dating, interpreting, and documenting sites and curating these artifacts to allow contemporary and future archaeologists access to primary scientific data. The initial reason for limiting collection, and thereby requiring all analyses to be performed in the field, was the assertion that collecting artifacts constituted an adverse impact on sites. In time, government agencies came to realize a second rationale for limiting collection: cost savings. Some Native American tribes, also, are opposed to artifact collection, making limited-collection and no-collection policies easier to justify.

Under federal law, collections made as a result of federally funded projects need to be cared for in perpetuity. Curation can be expensive, and the funding and space needed for curation are limited throughout the country. The driving force behind in-field analysis quickly became, and has continued to be, the minimization of curation responsibilities and expenditures.

Most often, in-field analysis is conducted under less-than-optimal conditions by field technicians who have little specific training in artifact analysis. Typically, in-field analysis is performed during survey, and the results are used to manage inventoried sites and to evaluate sites for eligibility for listing in the NRHP. Recently, some agencies have begun to entertain the possibility of using in-field analysis during data recovery for materials considered to have little research value (e.g., machine-made glass, metal fragments, and large ground stone artifacts). If artifacts are not collected during data recovery, in-field analysis results and other on-site documentation could be the only information that remains about a site, particularly if the site is destroyed during or subsequent to mitigation.

Limited-collection and no-collection policies are based on the assumption that in-field artifact analysis is comparable in quality to artifact analysis conducted in the laboratory. Further, implicit in limited-collection and no-collection policies are that contemporary and future archaeologists will accept the results of in-field analysis without question and that future research and management decisions will not need access to artifacts. The DoD and other federal agencies have expended, at a minimum, tens of millions of dollars on surveys based on these assumptions, yet no one has actually tested whether they are valid. Despite the heavy reliance on in-field analysis for inventory and evaluation and the considerations of performing in-field analysis during data recovery (see Chapter 1), very little is known about how accurate in-field analysis really is or how adequate it is for site interpretation.

Experimental Approach

The experiment presented in this report was designed to assess the adequacy and accuracy of in-field and digital-photograph artifact analyses at two prehistoric archaeological sites located on military installations in the western United States. One case study was the Soldier Creek site (AZ EE:7:164 [AMS]), at Fort Huachuca, in southeastern Arizona. The other site was FB 9583, in south-central New Mexico, on the East McGregor Range of the Fort Bliss Military Reservation. At each of these sites, samples of individually numbered ceramic and lithic artifacts derived from surface contexts were analyzed by two separate field technicians. The artifacts were then collected and analyzed by trained specialists who performed their analyses in a laboratory setting, using either the physical artifacts or, alternatively, only digital photographs of the artifacts. All analysts used the same typological system for identifying artifacts and also used field manuals that described the attributes of specific artifact types identified during the project.

Artifact identifications resulting from each of the analyses were entered into a relational database system using standardized terms and quality controls. The data from the analyses were compared between and among analyses to assess the levels of agreement between analyses and the adequacy of site interpretations that could be derived from each analysis. These assessments, which were performed separately for ceramic and lithic artifacts, are presented in Chapters 3 and 4 of this report.

Results

Several methods were used to assess analysis results. One method was to calculate an agreement index representing the proportion of observations that were in agreement with each other for each pair of analyses. For ceramic artifacts, agreement indexes were calculated for vessel element, ceramic ware, and ceramic type. For lithic artifacts, agreement indexes were calculated for production method, material type, and technological type. In assessing the agreement-index results, no assumptions were made regarding which set of observations (in-field, digital-photographic identification, or hands-on laboratory identification) was more or less likely to be accurate. Rather, the assessments were geared toward testing the levels of agreement between different sets of observations, regardless of level of expertise or the context in which artifacts were analyzed.

A striking result of the agreement-index assessments was that the levels of agreement between analyses were high only for the most general of artifact attributes (e.g., vessel element or production method). Agreement tended to be lowest for attributes with the greatest number of potential values (Tables 48 and 49). Variation in the levels of agreement among analyses, as indicated by the coefficient of variation, was typically highest for attributes with the greatest number of potential states. For vessel element and production method (attributes that had limited numbers of possible states), the average agreement indexes ranged from 0.89 to 0.97. For ceramic type and lithic-technological type (attributes that had large numbers of possible states), the average agreement indexes were much lower, ranging from 0.48 to 0.57. With the exception of material type for lithic artifacts, the average levels of agreement tended to be similar between the two sites for the same attributes (e.g., ceramic ware).

Figures 6 and 7 show the minimum, maximum, and average agreement indexes at the two sites for ceramic and lithic analyses, respectively. A downward trend in agreement as the number of types observed increased is evident in both figures. The precipitous drop in the average and minimum agreement that occurred between the type counts of 10 and 15, as illustrated in Figure 7, is due to the anomalous material-type observations made by one of the field technicians at Fort Huachuca, whose material-type identifications agreed with those of other observers for only around 10 percent of artifacts analyzed (see Chapter 4). The general trend suggests that we can only expect artifact analyses to agree, on average, around 50 percent of the time when specific artifact types commonly used to interpret sites are identified, such as specific ceramic types that can be used to interpret cultural or temporal affiliation or specific kinds of flaked stone tools that may be used to infer some of the kinds of activities performed at a site. For lithic-technological types, agreement ranged from 31 to 67 percent. For ceramic types, agreement ranged from 39 to 75 percent.

A second approach used to assess the analysis results was to test the accuracy and precision of observations made in the field and observations made using only digital photographs. For these assessments, the hands-on-laboratory-analysis results were treated as the "gold standard" for the project. Then, the levels of Type I and Type II errors in artifact identifications were estimated by calculating the false-negative and false-positive rates for each artifact type and site. In statistics, a false negative is a case in which

Table 48. Agreement Statistics for Ceramic-Artifact Analysis, by Site and Observational Category

Observational Category, by Site Location	Total No. of Types Observed	Agreement Index			
		Maximum	Minimum	Average	Coefficient of Variation
Fort Bliss					
Vessel element	2	0.95	0.80	0.89	6.6
Ceramic ware	8	0.85	0.60	0.71	12.2
Ceramic type	18	0.75	0.45	0.57	18.2
Fort Huachuca					
Vessel element	3	0.97	0.89	0.93	3.1
Ceramic ware	12	0.82	0.62	0.73	10.1
Ceramic type	30	0.58	0.39	0.49	12.4

Table 49. Agreement Statistics for Lithic-Artifact Analysis, by Site and Observational Category

Observational Category, by Site Location	Total No. of Types Observed	Agreement Index			
		Maximum	Minimum	Average	Coefficient of Variation
Fort Bliss					
Production method	4	0.98	0.96	0.97	0.8
Material type	9	0.80	0.59	0.67	12.0
Technological type	23	0.54	0.39	0.48	10.7
Fort Huachuca					
Production method	3	0.98	0.95	0.96	1.3
Material type	12	0.83	0.09	0.44	85.9
Technological type	19	0.67	0.31	0.49	34.9

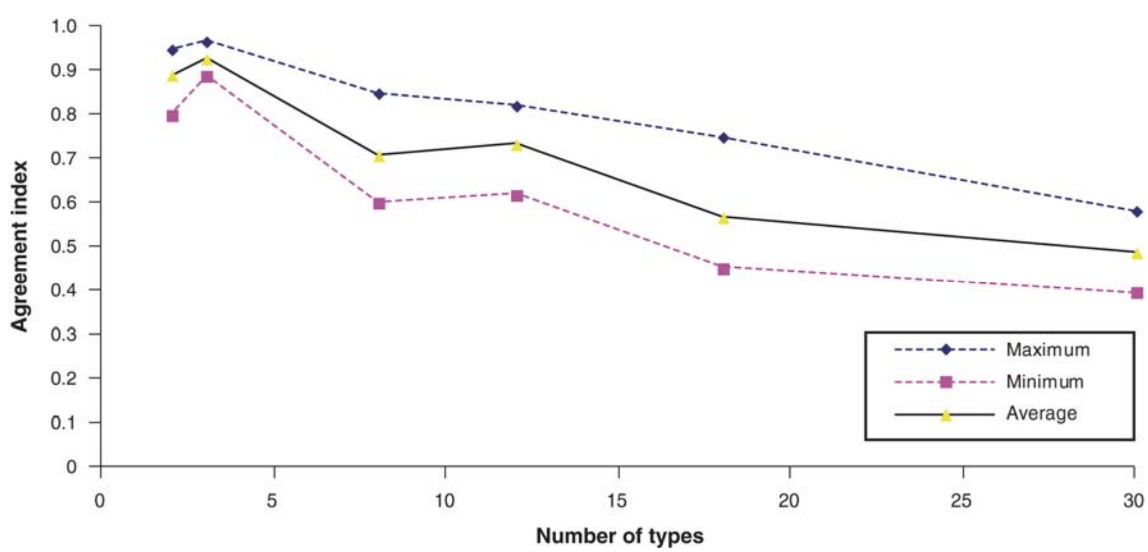

Figure 6. Plot of the minimum, maximum, and average agreement indexes for the ceramic-artifact analyses, according to the number of observed types.

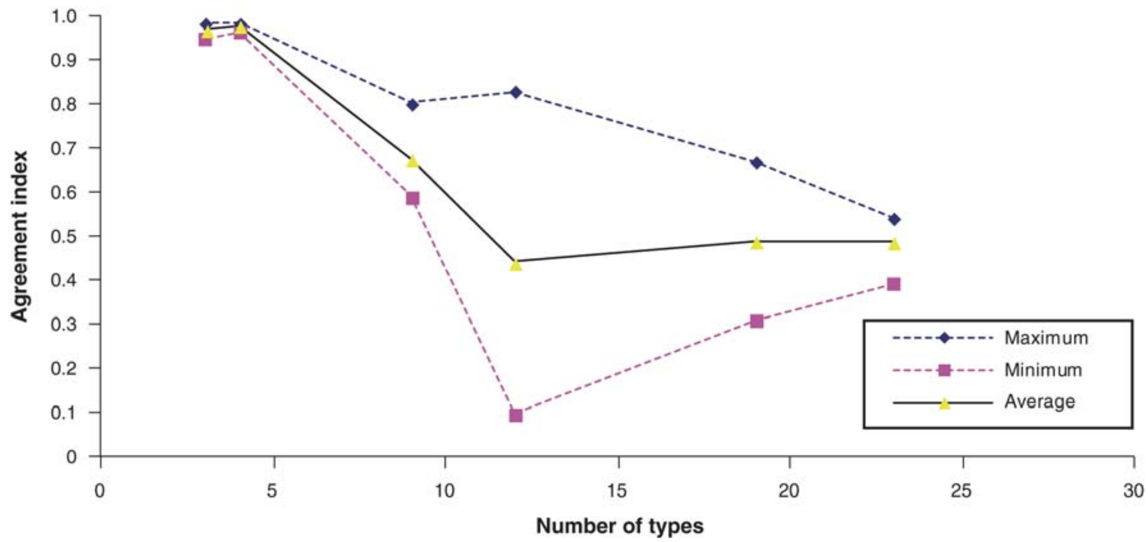

Figure 7. Plot of the minimum, maximum, and average agreement indexes for the lithic-artifact analyses, according to the number of observed types.

something is asserted to be absent when it is, in fact, present. In the context of this project, a false negative represented a case in which the hands-on analysis identified an artifact as a specific type, such as a biface, and one or more of the other analyses failed to identify the same artifact as a biface and instead identified the artifact as something other than a biface. A false positive is a case in which something is asserted to be present when it is, in fact, absent. In the context of this project, a false positive represented a case in which an artifact was identified by one or more analyses as a specific type, and the hands-on laboratory analysis identified the artifact as a different type. For instance, if an artifact was identified by the hands-on laboratory analysis as a flake, but one of the other analyses identified it as a biface, this would represent a false-positive case for the artifact type of biface.

On average, false-negative and false-positive rates calculated using the analysis results increased as the number of identifiable types increased (Figures 8 and 9). False-positive and false-negative rates were low and similar for general observations having to do with production method or vessel element but were larger and more disparate for more-specific and more-detailed artifact-type identifications, such as the identification of specific ceramic types or lithic-technological types. For ceramic types or lithic-technological types, false-negative rates approached 1 for all but the most-common or most-distinctive types (such as plain ware or projectile points, respectively); rare artifact types were often misidentified. In other words, many artifact types that could be used for site interpretation were misclassified most or all of the time. Only the most-common and most-recognizable artifact types were classified correctly most of the time.

False-positive rates either tended to approach 0, because a type was rarely identified outside the hands-on laboratory analysis (resulting in few or no false positives), or, alternatively, were high, because analysts frequently identified a ceramic artifact erroneously as one type when it was, in fact, another type. For instance, general artifact types, such as indeterminate painted ware or indeterminate red-on-buff, were often applied during the in-field or digital-photograph analysis to artifacts identified by the hands-on analysis as more-specific types, such as Encinas Red-on-brown. The general types (e.g., indeterminate red-on-buff), therefore, had high false-positive rates, because they were identified more often than they actually occurred. More-specific types often had low false-positive rates (e.g., Encinas Red-on-brown), because the types were identified by hands-on laboratory analysis but were rarely or never identified by in-field or digital-photograph analysis. In-field analysis tended to identify artifacts less precisely according to type than the hands-on laboratory analysis; digital-photograph analysis tended to identify artifacts more precisely, but according to types that were not identified by the hands-on laboratory analysis.

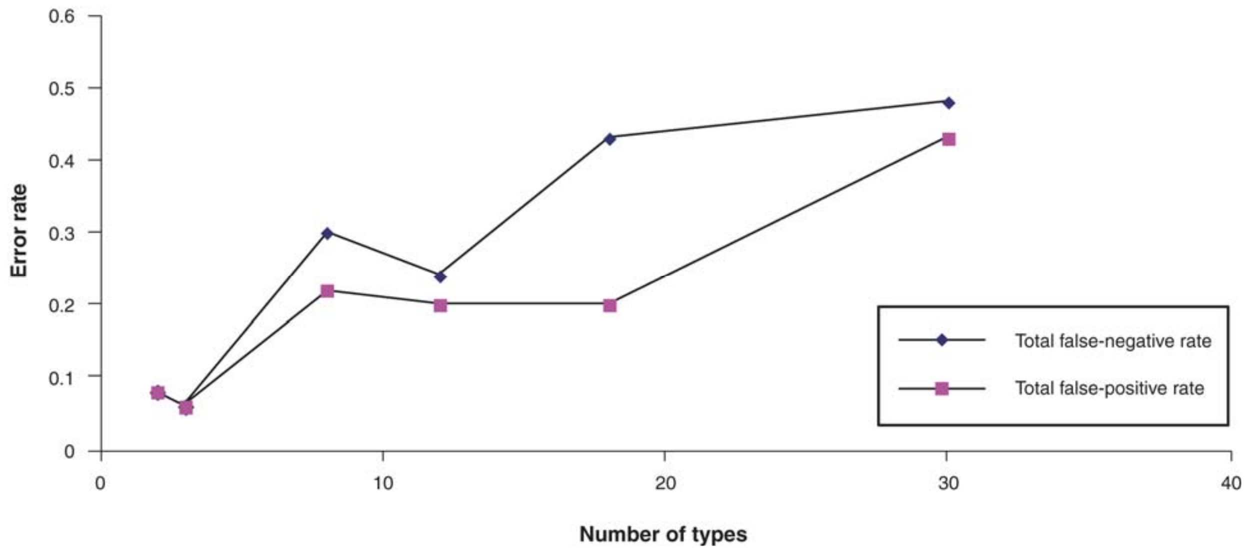

Figure 8. False-negative and false-positive error rates for the ceramic-artifact analyses, according to the number of observed types.

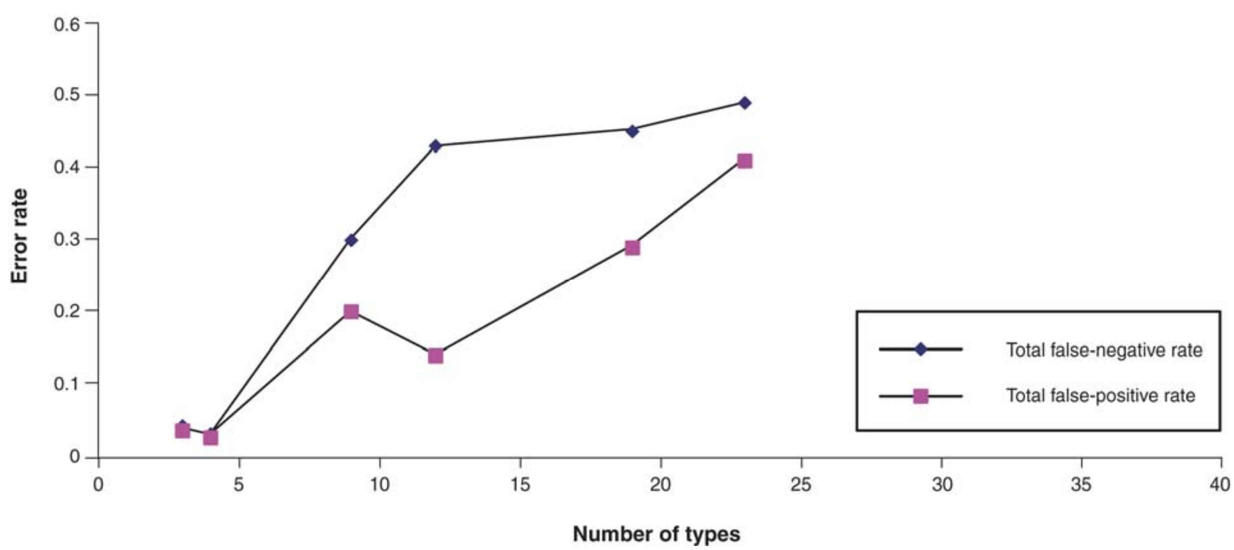

Figure 9. False-negative and false-positive error rates for the lithic-artifact analyses, according to the number of observed types.

Another method for assessing the accuracy of the in-field and digital-photograph analyses was to calculate two additional statistics for some of the more-common artifact types: Cohen's kappa and McNemar's test for bias. Cohen's kappa uses contingency tables to calculate the strength of agreement between observers based on the observed and expected levels of agreement. McNemar's test for bias uses the same contingency table to identify whether systematic bias may have occurred in artifact identification. Cohen's kappa showed that even for the more-common types present in a collection, the strength of agreement between the hands-on laboratory analysis and another analysis was often less than adequate. McNemar's test for bias showed that evidence for systematic bias was common.

For ceramic artifacts at Fort Bliss and Fort Huachuca, the strength of agreement was adequate for vessel elements for all analyses but was not often adequate for specific wares and types. At Fort Bliss, only the digital-photograph analysis was adequate for the most-common ceramic ware and type; in-field analysis was less than adequate in all cases. Evidence for systematic bias was not identified in the analysis of ceramic artifacts at Fort Bliss. At Fort Huachuca, the strength of agreement was adequate for all analyses only for vessel element and for plain ware ceramic artifacts. Otherwise, the strength of agreement for the most-common wares and types at Fort Huachuca was less than adequate for one or more analyses. In many cases, there was also evidence for systematic bias in artifact identification. Moreover, none of the in-field or digital-photograph analyses was universally better at identifying artifacts. Typically, one was better than another at identifying one or more types, but not in identifying all or a majority of important types.

For lithic artifacts at Fort Bliss, the strength of agreement was adequate for all analyses for the production methods of flaked stone and ground/battered stone, for the material type of sandstone, and for the technological type of cores. The strength of agreement for artifacts falling within the general categories of mano, metate, or retouched tool was also adequate. However, for specific kinds of retouched tools—biface, retouched piece, scraper, and uniface—the strength of agreement was less than adequate for most analyses. Although sandstone tended to be identified correctly, the strength of agreement for the more commonly occurring material types of chert and limestone was inadequate for several analyses. There was also evidence for systematic bias in the identification of common artifact types by individual analyses.

For lithic artifacts at Fort Huachuca, the strength of agreement was adequate for all analyses only for the production method of flaked stone. The strength of agreement was less than adequate for at least one analysis for all of the most-common material and technological types, including the more-general technological class of retouched tools. There was also evidence for systematic bias in the identification of a number of common artifact types. In part, these results suffered from the especially high error rates of one in-field analysis (see Chapter 4); the other two analyses tended to perform better and demonstrated an adequate strength of agreement for many, but not all, of the more-common material and technological types.

In addition to the accuracy of the in-field and digital-photograph analyses, the adequacy of results was also assessed, by evaluating the overall distribution of analysis results in order to determine whether differences in site interpretation could potentially result from the different analyses. In general, these assessments showed that there tended to be overarching similarities in the distribution of results and that many of the same types were identified in broadly similar numbers among analyses. However, it is often the case that artifacts that are diagnostic of specific cultural or temporal affiliations or of specific activities at a site are rare, and rare types were often misidentified. In-field analysis tended to identify artifacts according to more-general and less-precise types than hands-on laboratory analysis or digital-photograph analysis; also, both in-field analysis and digital-photograph analysis tended to identify types that were not present in the collections. Digital-photograph analysis and hands-on laboratory analysis tended to identify greater numbers of temporal components and more-precisely defined temporal components than in-field analysis. In contrast, in-field analysis tended to identify types that corresponded to broader periods than the types identified by hands-on laboratory analysis.

The results of the in-field and digital-photograph analyses also demonstrated the potential for such analyses to result in erroneous and incomplete interpretations of cultural and temporal affiliation. Artifact richness and the number of vessels that could potentially be inferred based on analysis results tended to be greater for hands-on laboratory analysis or digital-photograph analysis, suggesting that the intensity and diversity of activities interpreted to have occurred at a site could be substantially lower than if the results of the in-field analysis were used for interpretation. Also, in-field and digital-photograph analyses tended to identify more flaked stone tools than were indicated by the hands-on laboratory analysis (although the identification of fewer tools occurred in one case). At Fort Huachuca, the hands-on laboratory analysis identified a total of 10 flaked stone tools, whereas the digital-photograph analysis and one in-field analysis identified more than 20 flaked stone tools each, and the other in-field analysis identified only 4 flaked stone tools (see Table 44).

All together, these results suggest that accurate site interpretation based on in-field analysis or digital-photograph analysis is possible but not necessarily probable. Some analyses will perform well for some artifact types but will not perform well for other artifact types. Other analyses will perform poorly for many artifact types but can also perform fairly well for other artifact types. In some cases, digital-photograph analysis performed better than in-field analysis, but in other cases, the opposite occurred. There also tended to be variation between in-field analyses in the accuracy of artifact identifications; one in-field analysis could be fairly accurate for many, but not all, artifact identifications, and the other was inaccurate in many cases. Rare and important artifacts were missed most often by in-field analysis, whereas digital-photograph analysis identified rare and important artifact types, but not necessarily the right ones.

Moreover, assessment of analysis results shows that some important inferences based on in-field or digital-photograph analysis alone are likely to be erroneous or incomplete. For instance, artifact types that might be used to infer specific activities, such as the tabular knives identified by one field technician at Fort Bliss (see Chapter 4), may be identified by in-field analysis when they are not actually present in a collection. In such a case, they could lead to a manager's conclusion that agave procurement was performed at a site when, in fact, it may not have been (Doolittle and Neely 2004). Conversely, rare and important material types, such as obsidian, may be missed by both in-field and digital-photograph analyses (as happened at Fort Bliss), or one analysis may systematically identify the wrong material types for many artifacts (as happened at Fort Huachuca). In the former case, important information that could be derived from obsidian artifacts would be lost (see Haury 1976; Hoffman 1997; Liritzis and Diakostamatiou 2002; Liritzis and Stevenson 2012; Riciputi et al. 2002; Shackley 2005). In the latter case, an investigator might conclude that the users of a site had different material preferences or accessed different quarry areas than they actually did.

Also, based on in-field analysis, an investigator might conclude that fewer activities occurred at a site than was the case, because of artificially low artifact richness; that ceramic-vessel use was less intense than was the case, because of erroneously low vessel counts; or that flaked-stone-tool use was either more or less intense than was the case, because of erroneously high or low flaked-stone-tool counts. Different kinds of ground stone or flaked stone technology could also be inferred based on the results of the in-field or digital-photograph analysis. For instance, one analysis might identify a substantially greater number of two-handed manos in a collection than were likely present, leading to the potential inference that intensive processing of maize occurred at a site (Adams 2002), whereas another might identify several basin metates and milling stones that were not likely present at a site, leading to the inference that grinding activities focused on the use of less-formal tools. Fewer cultural or temporal affiliations and less-precise chronologies could potentially be inferred from the results of in-field analysis. On the other hand, although digital-photograph analysis appears to have the potential to achieve a more-accurate and more-precise picture of a site's temporal components than in-field analysis, it can still miss rare components.

The Accuracy and Adequacy of In-field Analysis

Accuracy has two components: one is empirical, and one is inferential. Archaeological investigations that rely solely on in-field analysis have to demonstrate that a sherd classified in the field as Rincon Red-on-brown really is Rincon Red-on-brown and not Rillito Red-on-brown, or they must provide a means by which other archaeologists can verify the classification. As shown in the current study, digital photographs may not be adequate to enable verification of the in-field classification. The problem is not simply of accuracy but of adequacy of interpretation. For example, an archaeologist might correctly identify a sherd as Hohokam Buff Ware, but tremendous information is lost if the sherd could have been further typed as Sacaton Red-on-buff. The current study showed that many sherds at Fort Huachuca, for instance, were identified in the field as relatively general types and that more-precise types were identified primarily by trained specialists in a laboratory setting. This problem is endemic at the survey level. Most field-crew members are not adequately trained, many artifacts cannot be adequately cleaned, and field conditions, particularly in the summer, are not conducive to artifact analysis.

An example of the pernicious effect of the no-collection policy on archaeological research can be drawn from research conducted in the Vekol Valley and Sand Tank Mountains of the western portion of the

Papaguería. Between 1989 and 1994, SRI conducted three major surveys in these areas for the U.S. Air National Guard Air Force Reserve. Most of the 100+ sites found were artifact scatters with very little likelihood of subsurface deposits. Surface artifacts were collected as part of the 1989 survey but not during the subsequent two surveys.

More recently, the results of the surveys were reanalyzed and published in an article on the settlement and subsistence of the region that was based on the survey results. The article argued for the existence of two Hohokam communities and a much more complex cultural setting than had been previously interpreted for the resources (Altschul et al. 2008). Unfortunately, the article was ultimately only able to use data from the first survey, during which artifacts had been collected—not because the data from the latter two surveys were faulty, but because without detailed laboratory analysis of the collected artifacts, the sites from the second and third survey could only be classified very simply, according to the cultural affiliations of Hohokam, Patayan, or "other." They could not be accurately classified according to distinctive periods corresponding to Hohokam and Patayan cultural histories (e.g., Pioneer, pre-Classic, or Classic period Hohokam or Patayan I, II, or III). Lots of time, money, and effort were expended to document these sites, but the returns from those cases in which no collections had been made were minimal. Had this been data recovery and had the sites been subsequently destroyed without any collections made, the losses would have been irretrievable.

In the digital age, one might argue that it is possible to document surface artifacts sufficiently, such that we do not need to keep the physical objects. The truth of such a conclusion, however, is still very much in question. The current study showed that digital-photograph analysis, even when conducted by a trained specialist, can be just as inaccurate as, and sometimes less accurate than, in-field analysis. Although digital-photograph analysis conducted by a trained specialist could produce more-refined results, it could also result in inaccurate and misleading site interpretations. At both Fort Huachuca and Fort Bliss, digital-photograph analysis resulted in the identification of multiple distinctive artifact types (such as specific ceramic types or ground stone tools) that were not likely to have been present in the collection and that could have led to erroneous conclusions about the cultural and temporal affiliations of the sites and about specific kinds of activities performed there.

In addition, digital curation is certainly not without cost. Preparation of metadata, transferring of files to archival formats, and migration of software and hardware may, in the end, be just as expensive as curation of the physical objects. Moreover, in order to curate digitally, the DoD would have to guarantee the long-term storage and preservation of digital materials.

Clearly, analysis of artifacts impacts the interpretation of a site from an archaeological standpoint. One might ask, does misidentification of artifacts also impact the management of a site? If so, how? Conversely, if all that is needed to manage a site are only the most basic data on site location, size, and content, then what is the purpose of collecting any specific information about artifact types or attributes, particularly if those data are routinely insufficient for accurate identification and cannot be trusted? Following this logic, it could be argued that efforts expended to analyze artifacts in the field are not well spent and should be placed elsewhere. Why analyze artifacts at all? In some cases, having detailed and accurate information about a site will not affect management decisions, but only when minimal information is needed to understand impacts to a site and when adverse effects to a site will always be avoided. Avoidance is increasingly difficult as more land area is devoted to training activities, and as greater numbers of sites are discovered, the options for avoidance become limited.

NPS guidelines specify that evaluation of the eligibility of a property has to be based on careful consideration of the significance and integrity of a site (Little et al. 2000; Parker and King 1998; Shrimpton 2002). The significance of a site is supposed to be determined based on the relevance of a property to a historic context. Properties considered eligible for listing in the NRHP under Criterion d are significant based on their relevance to important research questions or data gaps. The determination of significance is thus based on issues that speak to the importance of a property to U.S. history and cultural heritage and are, in most cases, inherently of archaeological, historical, anthropological, or social concern. Having accurate and reliable information about a site is crucial to interpreting its significance and to determining whether it has sufficient integrity to convey its significance and to what degree and in what ways adverse effects could impact that significance. In other words, understanding whether and how a site will be impacted by a federal undertaking and how those impacts should be resolved is partly dependent upon having accurate information about the attributes of a site, as determined through artifact analysis. With inaccurate data, sites may need to be visited multiple times to gain the information needed for site evaluation and identification of impacts, greatly expanding the effort needed to complete evaluations and to plan projects. With artifact collection and laboratory analysis, the level of effort needed for evaluation would arguably be less. Moreover, the artifacts that support these determinations could, themselves, be reviewed, and interpretation revised or validated, without costly fieldwork, even for cases in which long periods of time have passed since inventory was conducted.

Much of the information needed to reliably identify the current or long-term research potential of a site, in order to evaluate its significance under Section 106 and support the stewardship responsibilities mandated under Section 110 of the NHPA, cannot be developed without accurate artifact information. Identification of the temporal and cultural affiliations of sites, as well as site function, is routinely established through artifact analysis. Indeed, several of the variables that many investigators have indicated as crucial to establishing representative samples of sites for long-term preservation and for determining the long-term data potential of sites are derived primarily from artifact analysis: temporal period, site function, assemblage diversity, and cultural affiliation (Briuer and Mathers 1997; Sebastian 2009). The analysis presented in this volume shows that all of those variables could be inaccurately assessed, sometimes grossly so, when based on in-field or digital-photograph analysis.

Another reason that accurate artifact identification is important to management is that consultation regarding the eligibility and treatment of historic properties is a fundamental component of the compliance process mandated by Section 106 of the NHPA. Adverse effects to historic properties that are eligible for listing in the NRHP need to be considered *in consultation* with other parties, including SHPOs, Tribal Historic Preservation Offices, Native American tribes, and, as necessary, the ACHP. Thus, engaging stakeholders in the identification and consideration of adverse effects and establishing a sense of trust among consulting parties should be important goals of DoD CRM programs. Establishing trust and good working relationships is crucial to ensuring that the process runs smoothly and to demonstrating that the DoD is performing its compliance responsibilities in good faith. Having accurate data that can be counted on for consultation regarding adverse effects to historic properties is one means of establishing trust. This experiment makes clear that site interpretation based on in-field analysis can be inaccurate and that rare and important site components may be missed or misinterpreted because of faulty artifact data. If these possibilities become accepted truths, then other, more basic and important data, such as data on location and integrity, may also not be trusted. Lacking confidence, installations may never get buy-in from stakeholders and could either fail to successfully complete the compliance process or be required to engage in other, unwelcome and costly efforts, such as dispute resolution and termination of consultation.

What to Collect? That is the Question

Questions of what to collect, how many artifacts of each material class need to be analyzed, and how many, if any, collected artifacts need to be curated in perpetuity have been debated by archaeologists for generations, and archaeologists have traditionally been guided by two principles in answering them. The first of these can be called the stewardship principle. Because archaeology is a destructive act, archaeologists who excavate a site have a responsibility to ensure that the site is adequately documented, such that future archaeologists, Native Americans, and other members of the public can use the data obtained to continue to learn about the past. As Childs (1995:n.p.) has noted, “the reality is that, once a site is excavated, these materials are often the only remaining evidence of a past culture. Not surprisingly, they are proving increasingly valuable for thesis, dissertation, and other research projects.” Stewardship means that artifacts are properly identified, cataloged, and curated, along with notes, databases, photographs, maps, and all other materials necessary to understand a site.

The second principle that provides a rationale for artifact collection can be called the science principle. Archaeologists, like all scientists, forward arguments based on their interpretations of data. Unless an archaeologist allows his or her peers, as well as future generations of archaeologists, to confirm or refute his or her interpretations by reexamining the basic data upon which a claim is based, such arguments cannot be verified; hence, they are considered suspect, if they are not simply rejected out of hand. For an article published in *Science*, Michael Bawaya (2007) interviewed Terry Childs (then a curator with the NPS in Washington, D.C., and currently a curator with the USDI), who provided an example of one of the pitfalls of in-field analysis: the inability to reevaluate materials used to interpret a site. Childs noted that in one excavation in Maryland, a colleague left the artifacts from a historical-period house in the ground rather than cleaning and analyzing them in the laboratory. Based on the in-field analysis, the house was interpreted as dating to the twentieth century. However, more-detailed subsequent excavation revealed that the house dated instead to the nineteenth century (Bawaya 2007:1025). No-collection surveys, which are on the rise, result in similar problems. Interviewed for the same article, Christopher Pullium, an archaeologist with the USACE in St. Louis, Missouri, cautioned that the integrity of archaeology as a scientific discipline is undermined without collection: "If one can't replicate research results or reanalyze the materials from a site, then [archaeology] can't proclaim to be a science" (Bawaya 2007:1026).

A decision to allow in-field analysis in lieu of collection must demonstrate how archaeologists will still be able to meet their scientific and stewardship responsibilities. Because curation is expensive and curatorial space is at a premium, all archaeologists must be cognizant of the crisis and ensure that collections are prudent and reasonable. One potential way to address the situation is to collect a sample of artifacts from a site that can be analyzed in a laboratory setting by a trained specialist. Even this solution, though practical, has some potential drawbacks that should be considered, based on the results of this study (see also Graesch 2009).

For instance, Graesch (2009) showed that the recovery of archaeological materials is highly dependent on field-crew experience and recovery method. Graesch's (2009) study documented that recovery rates for artifacts in the field can be as low as less than 10 percent (compared to the total number of artifacts recovered after resifting field-sifted sediment in a laboratory setting). Artifact-recovery rates for field crews with limited experience can be both low and highly variable, ranging from 5.8 to 33.2 percent. Recovery rates for highly experienced crews were much higher and less variable, ranging from 50.5 to 61.5 percent. Overall, the average recovery rate for crews with a variety of levels of experience was 47.5 percent. Graesch's (2009) analysis showed that even among highly experienced crews, the percentages of artifacts of common types that were missed in the field ranged from 43 to 62 percent, and the percentages of artifacts of rare types that were missed in the field ranged from 51 to 100 percent. In terms of interpretation, field-recovered artifacts were able "to discern major inter-household differences in terms of emphases on particular activities [but were] overall poor indicators of the intensity at which these activities were performed" (Graesch 2009:774). Thus, Graesch (2009:774) argued that field data resulting from the study could not be used "to detect subtle but important differences in the focus and intensity of household activities." Ultimately, Graesch (2009:777) concluded that the "recovery of artifact assemblages that accurately reflect the variety and abundance of artifacts occurring in excavated deposits hinges on the retention of a representative portion of screen residue for laboratory analysis."

Collecting a representative sample according to type could be unreliable, simply because the identification of artifact types by in-field analysis is unreliable. As was shown in Chapter 3, many ceramic types failed to be identified correctly in the field. So, if field crews were instructed to collect all painted ceramic artifacts or a sample of painted ceramic artifacts, a substantial proportion of painted ceramics at a site could be erroneously identified in the field as unpainted, or vice versa. This certainly could have happened at the Fort Huachuca site, for instance.

Rather than collect all or a sample of painted ceramic artifacts, a conservative approach that has greater potential to preserve important information about a site would be to collect a sample of decorated ceramic artifacts and a sample of undecorated ceramic artifacts. Even better would be to make these collections according to site component, locus, or feature type. Such an approach would have a better chance of resulting in a representative sample, by lessening the impact of potential artifact-identification errors.

The same goes for lithic tools, another artifact type that is commonly considered diagnostic of site function. If a field crew were directed to collect all or a sample of lithic tools, it is likely that the collection would include other lithic artifacts not classified as tools and that important lithic tools would be missed. Thus, it might make sense to have crews collect all lithic tools up to a certain number and to collect a representative sample of cores and flakes, because laboratory analysis could reveal whether some of the artifacts identified as tools in the field were actually cores or flakes, or vice versa.

The collection of some artifacts, such as some ground stone artifacts, could be difficult to justify in situations of limited curation space and funding, because of their bulk. These kinds of artifacts may require more-thorough documentation and measurement in the field to provide enough information for the artifacts to be interpreted more fully by a trained specialist or reevaluated in the future.

Recommendations for Best Practices

This report would not be complete without providing recommendations for best practices. Given that this study has shown that in-field and digital-photograph analyses of many artifact classes and types were often inaccurate or imprecise, it is not possible at this time to suggest which artifact types or classes are likely to be identified correctly by in-field or digital-photograph analysis. However, recommendations for improving the training of in-field analysts, the tools available for in-field analysis, and the conditions under which in-field analysis is conducted can be provided. Recommendations can also be provided regarding further assessment of in-field-analysis results. Below is a series of recommendations for best practices.

- *Provide field technicians with periodic training in artifact analysis that is specific to particular subregions, survey areas, and artifact classes*. Field technicians can perform well in identifying some artifact types, but they may often not have adequate experience to identify unusual or rare artifact types. When in-field analysis has to be performed by field technicians, provision of training will at least give field technicians some of the tools they need to accurately identify artifacts according to type. Workshops in artifact analysis could periodically be offered by the DoD or other agencies to provide such training. Review of artifact types likely to be encountered in a project area prior to fieldwork would also improve the quality of in-field analysis. Review could be accomplished prior to fieldwork by having a trained specialist review existing collections with the field crew or by visiting one or more sites with artifact types likely to be encountered during fieldwork.
- *Develop technical field manuals pertinent to specific regions and subregions that can be used to present standardized and explicit information on how to identify the artifact types likely to be encountered at archaeological sites in a given region*. Technical field manuals that provide illustrations of artifact types along with descriptions of their specific attributes will provide a common typological system and a common information base for in-field artifact analysis. Such field manuals can provide information concerning what artifact attributes to look for, how to identify them, and which attributes are important to measure and record. The use of field manuals not only can improve the accuracy and reliability of in-field artifact analysis but also can contribute to ensuring that data obtained by different crews and organizations are comparable across surveys (see Railey 2010).
- *Test the accuracy and adequacy of in-field artifact analyses performed by individuals tasked with conducting in-field analysis for archaeological projects.* Field-crew members who are slated to

perform in-field analysis, particularly for large projects, can be tested in order to rate their ability to identify artifacts of different artifact classes and types. Those who demonstrate the greatest ability in identifying specific artifact classes can be tasked with analyzing those artifact classes in the field. Deficiencies in artifact identification can also be identified through testing, in order to identify where training efforts need to be placed as well as cases in which artifact-analysis results are likely to be unreliable. It may also be possible to develop a rating system that can be used to identify field-crew members with demonstrated abilities to perform in-field artifact analysis.

- *Test the accuracy and adequacy of in-field artifact analysis performed during previous survey.* In order to assess whether previous survey resulted in reliable artifact data, previously surveyed sites subjected to in-field analysis can be revisited by trained specialists in order to verify artifact-analysis results. In some cases, it may not be possible to identify the specific artifacts analyzed during fieldwork, but it may be possible to assess whether important types were missed and which types were likely to have been misclassified. It may also be possible, in some cases, to reanalyze collections from previous surveys. The results of laboratory analysis of collections could be compared with artifact identifications made in the field. If artifacts were individually numbered, the same quantitative methods used in this report to assess analysis results could be applied to collected artifacts. Such efforts would allow managers to estimate the level of confidence that could be placed in the results of previous survey efforts and to identify sites or survey areas that may need to be revisited. These efforts could also be combined with eligibility evaluations.
- *When in-field analysis is to be conducted, have trained specialists perform the analysis during inventory of large and important sites, during eligibility evaluations or site revisits, and during data recovery.* For some projects, known sites are rerecorded during survey or are revisited to perform eligibility evaluations. Particularly if such sites have components that are known or suspected to be especially important or unusual, having a trained specialist perform the analysis will likely provide more-accurate and more-complete artifact-analysis results. Because data recovery could result in the destruction of a site or site components, it is imperative that trained specialists perform any in-field analysis that is conducted, because in such cases, in-field analysis may constitute the only remaining record of a site's artifact content.
- *Collect a representative sample of portable artifacts from sites subjected to in-field analysis.* In many cases, collecting a representative sample of artifacts from a site will involve the collection of small numbers of artifacts. Overall, this might amount to one or two archival boxes of artifact collections for an individual survey. Often, collected artifacts can be restricted to artifacts identified as diagnostic of temporal affiliation, cultural affiliation, or tool use. In addition to collecting examples of these artifacts, it would be useful, when possible, to collect samples of artifact types according to artifact class, such as flaked stone tools, reduction debris and nuclei, ceramic artifacts, and ground stone tools. Collection of a representative sample of material types would also be important, because accurate identification of material types can be difficult to achieve in the field. Collecting representative samples of artifacts according to artifact class and material type would help to reduce the potential for bias in sample selection (see above discussion).
- *Intensively document artifacts analyzed in the field.* In many cases, simply providing data on an artifact type will likely be insufficient for the identification to be validated or considered sufficient and reliable. Digital photographs also do not appear to consistently provide sufficient information for an accurate identification to be made. Key measurements and identification of morphological attributes or design elements could provide data that could be used to evaluate the accuracy of a field identification. These could be attributes specified in a technical field manual. Intensive photographic documentation of artifacts from multiple angles, using standardized background templates, could also preserve information about the artifacts analyzed in the field. Along with a metric scale, digital photographs could be taken using an industry-standard gray

card (such as an 18 percent gray card). This would enable digital photographs to be color-corrected in the laboratory using the known color values of the gray card. In rare instances, especially important artifacts at some sites could be documented using lidar or other three-dimensional scanning technologies. It is important to recognize that intensive documentation of artifacts in the field does not come without added cost. The value-added cost of additional documentation vs. the cost of artifact collection and curation should be assessed to determine whether in-field analysis substantially reduces the cost of a project.

- *Use handheld recording devices or personal digital assistants (PDAs) with standardized entry forms to conduct in-field analysis*. PDAs can be programmed with site forms that are populated from drop-down menus and lists that enable data to be entered according to a standardized format and typological system. Use of such devices and programs can reduce transcription errors, streamline data entry, and ensure that data entry is consistent between observations and observers. In some cases, it may also be possible to link artifact photographs directly to individual artifacts recorded in the field as well as to record locational information associated with the analyzed artifacts.
- *Where possible, analyze provenienced artifacts in a laboratory setting, and return them to their discovery locations after analysis*. The most-accurate analysis of artifacts will likely be achieved by a trained specialist in a laboratory setting. In some cases, it may be possible to temporarily remove artifacts from their discovery locations in order to analyze them in a laboratory, thereby providing accurate analysis while also minimizing collections. For large and intensive investigations, a temporary field laboratory can be established. Extra effort to point-provenience artifacts may need to be expended, if only to enable replacement of artifacts after analysis in locations that are close to their exact discovery locations in the field. Use of collection units for such a strategy instead of point-proveniencing would be more cost-effective but could result in artifacts' not getting replaced in their original positions. If minimizing impacts were a goal of a project, that goal might be undermined if collection units were used instead of point-proviencing. For cases in which artifacts need to be returned to their original discovery locations, photographic documentation and/or mapping of the areas from which artifacts have been temporarily collected would help to ensure that they are returned to locations that are close to their original discovery locations.
- *Engage stakeholders in survey methods and procedures, providing the pros and cons of in-field vs. laboratory analysis and collection vs. noncollection of artifacts*. It is often assumed by federal land managers that stakeholders, particularly Native American tribes and communities, would rather not have sites disturbed by artifact collection. However, if it is clear that the default method, in-field analysis, is likely to produce less-than-desirable results upon which the stakeholders, the SHPO, and the federal agency will be asked to base resource decisions, it is possible that these groups may opt to have sites minimally disturbed by limited, systematic collections. At the very least, as part of their consultation efforts, federal agencies need to provide stakeholders with assessments of the accuracy of the provided archaeological data upon which they are asked to base recommendations of significance, project effects, and project alternatives.

Where Do We Go From Here? Archaeological CRM Surveys in the Twenty-First Century

U.S. archaeology made a Faustian bargain in the late twentieth century: to cut costs and to alleviate perceived Native American concerns, archaeologists unilaterally eliminated or limited artifact collection on surveys. We cut our costs dramatically, but we have paid a steep price. Project areas surveyed 10 or

15 years ago routinely have to be resurveyed. Determinations of eligibility can be based on such unreliable data that stakeholders believe that they are considering one type of site when, in fact, the site under consideration is fundamentally different from what has been reported. Archaeologists and stakeholders alike are constantly surprised to find during data recovery that a carefully crafted and well-considered research design must be heavily revised or completely abandoned. In such cases, project sponsors are alarmed to learn that budgets for archaeology must be increased substantially in order to successfully complete a project. Many of these problems stem from the relegation of nearly all artifact analysis to in-field analysis, in the absence of collection, leading to decisions and management strategies based on faulty data that are not easily validated.

Voices are now being raised to question the correctness of limited-collection and no-collection policies. The USACE Mandatory Center of Expertise for the Curation and Management of Archaeological Collections, for example, has advised that a minimum representative scientific sample be collected during archaeological projects, including during survey, because samples are needed for some kinds of analysis and because in-field identifications cannot be scientifically verified when no collections are made. But is it too late to put the genie back in the bottle?

The decision to collect or not to collect on survey should not be prejudged. It falls to every archaeologist and cultural resource manager to make the best case for specific field and laboratory methods based on explicit assumptions for each particular project evaluated by representatives of federal agencies, SHPOs, tribes, and others. **Blindly following limited-collection or no-collection policies should not be viewed as a best practice.** DoD installations have a big stake in this debate. Certainly, CRM programs must be cost-effective, and it is tempting to do everything possible to cut expenditures.

However, saving money at the expense of our national heritage is a poor bargain. For nearly a century, the collection of artifacts has allowed for the continued investigation of the past. This information has been and continues to be of benefit to the many publics of the United States. But it comes at a cost. Artifacts and associated documentation need to be curated in perpetuity in accordance with federal collection-management guidelines and with the support of federal funds. Some artifacts can be deaccessioned, but the degree to which deaccessioning will reduce the burden and relieve the curation crisis is unknown. Archaeologists, installation managers, SHPOs, Native American tribes, and other interested parties need to weigh the options. What will be preferred: (a) collecting, analyzing, and curating artifacts in order to interpret significance, resolve effects, consult with stakeholders, and preserve the long-term data potential and scientific value of sites or (b) continued acquisition of, and dependence of DoD CRM decisions on, artifact-analysis data collected in the field, which are likely to be of questionable accuracy and reliability and of limited utility?

The discussion thus far has been in the abstract. How much is our past worth? What needs to be understood in much more explicit and much clearer detail is what the real costs of collection, laboratory analysis, and curation are, to the DoD and to the public, vs. the real costs of in-field analysis and limited collection or noncollection of artifacts. The assumption is that because collection and curation add lines to a budget, in-field analysis must be cheaper and more cost-effective than collection. Whether this is really true in the long-run is not known. What are the costs of revisiting sites and performing additional recording and analysis in support of evaluation efforts necessitated by inaccurate artifact data? What are the costs of investing more time in the in-field recording of artifacts and the curation of related documentation? What are the social, political, and cultural-heritage costs of losing resource value by not knowing enough about a resource to interpret its significance and decide on the course of stewardship?

Before we can make informed decisions about surface-artifact-collection or in-field-analysis strategies, we need to have more data. A follow-up study should be initiated to compare the costs of different approaches in the analysis of surface artifacts to the potential consequences of the various strategies. For example, in-field analysis might lead to a greater number of sites determined eligible for listing in the NRHP, thus requiring increased time and effort to manage them. We can model these management costs and compare them to the additional costs of analyses that would have alleviated the management costs. Once these cost comparisons have been made, the DoD will be in a better position to assess the costs in time and effort and to reach a decision as to the most effective way to manage our archaeological heritage.

REFERENCES CITED

Adams, Jenny L.
2002 *Ground Stone Analysis: A Technological Approach*. University of Utah Press, Salt Lake City.

Adams, Karen R.
2009 Flotation Analysis. In *Archaeological Investigations at Five Prehistoric Sites on the Eastern Flanks of the Organ Mountains, Fort Bliss, Doña Ana County, New Mexico*, edited by Christine G. Ward, Bradley J. Vierra, and Kari M. Schmidt, pp. 283–292. Historic and Natural Resources Report No. 07-53. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas. Technical Report 09-43. Statistical Research, El Paso.

Adams, William Y., and Ernest W. Adams
1991 *Archaeological Typology and Practical Reality: A Dialectical Approach to Artifact Classification and Sorting*. Cambridge University Press, New York.

Adler, Michael A. (editor)
1996 *The Prehistoric Pueblo World, A.D. 1150–1350*. University of Arizona Press, Tucson.

Altschul, Jeffrey H.
1986 Statistical Evaluation of the Collection Procedure. In *The Valencia Site Testing Project: Mapping, Intensive Surface Collecting, and Limited Trenching of a Hohokam Ballcourt Village in the Southern Tucson Basin*, by Mark D. Elson and William H. Doelle, pp. 25–30. Technical Report No. 86-6. Institute for American Research, Tucson.

Altschul, Jeffrey H., Christopher J. Doolittle, and Jeffrey A. Homburg
2008 The Archaeology and Prehistoric Culture of the Sand Tank and Vekol Valleys. In *Fragile Patterns: The Archaeology of the Western Papaguería*, edited by Jeffrey H. Altschul and Adrianne G. Rankin, pp. 253–267. SRI Press, Tucson.

Altschul, Jeffrey H., Rein Vanderpot, César A. Quijada, and Robert A. Heckman
forthcoming People of the Grassland: Archaic and Formative Cultures of the Upper and Middle San Pedro Valley. In *The Archaeology of a Land Between: Regional Dynamics in the Prehistory and History of Southeastern Arizona*, edited by Henry Wallace. University of Utah Press, Salt Lake City.

Anderies, John M., Ben A. Nelson, and Ann P. Kinzig
2008 Analyzing the Impact of Agave Cultivation on Famine Risk in Arid Pre-Hispanic Northern Mexico. *Human Ecology* 36(3):409–422.

Anderson, Sally
1993 Archaic Period Land Use in the Southern Tularosa Basin, New Mexico. In *Preliminary Investigations of the Archaic in the Region of Las Cruces, New Mexico*, edited by Richard S. MacNeish, pp. 48–67. Historic and Natural Resources Report No. 9. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Basehart, Harry W.
1971 Mescalero Apache Band Organization and Leadership. In *Apachean Culture and Ethnology,* edited by Keith H. Basso and Morris E. Opler, pp. 35–50. University of Arizona Press, Tucson.

Basso, Keith H.
1983 Western Apache. In *Southwest*, edited by Alfonso Ortiz, pp. 462–488. Handbook of North American Indians, vol. 10, William C. Sturtevant, general editor. Smithsonian Institution, Washington, D.C.

Bawaya, Michael
2007 Curation in Crisis. *Science* 317:1025–1026.

Beck, Charlotte, and George T. Jones
1989 Bias and Archaeological Classification. *American Antiquity* 54:244–262.

1994 On-Site Artifact Analysis as an Alternative to Collection. *American Antiquity* 59:304–315.

Beckes, Michael R., David S. Dibble, and Martha Doty Freeman (editors)
1977 The Cultural Resource Base. In *A Cultural Resource Inventory and Assessment of McGregor Guided Missile Range, Otero County, New Mexico*, *Vol. 1*. Research Report No. 65. Texas Archeological Survey, University of Texas, Austin. Historic and Natural Resources Report No. 75-03. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Beers, Yardley
1957 *Introduction to the Theory of Error*. Addison-Wesley, Reading, Massachusetts.

Bishop, Yvonne M. M., Stephen E. Fienberg, and Paul W. Holland
1975 *Discrete Multivariate Analysis: Theory and Practice*. Massachusetts Institute of Technology Press, Cambridge, Massachusetts.

Bohrer, Vorsila L.
2007 *Preceramic Subsistence in Two Rock Shelters in Fresnal Canyon, South Central New Mexico*. Archaeological Series No. 199. Arizona State Museum, University of Arizona, Tucson.

Boyd, C. Clifford, Jr.
1987 Interobserver Error in the Analysis of Nominal Attribute States: A Case Study. *Tennessee Anthropologist* 12(1):88–95.

Brennan, Paul, and Alan Silman
1992 Statistical Methods for Assessing Observer Variability in Clinical Measures. *BMJ* (304)6840:1491–1494.

Briuer, Frederick L., and Clay Mathers
1997 *Trends and Patterns in Cultural Resource Significance: An Historical Perspective and Annotated Bibliography*. Technical Report EL-97-5. U.S Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory, Champaign, Illinois.

Butler, William B.
1979 The No-Collection Strategy in Archaeology. *American Antiquity* 44:795–799.

Carlin, J.
1993 *Repeatability and Method Comparison: Introduction to Reliability and Agreement*. Clinical Epidemiology and Bio-Statistics Unit, Royal Children's Hospital Research Foundation and the Melbourne University Department of Paediatrics, Melbourne, Australia.

Carmichael, David
1986 *Archaeological Survey in the Southern Tularosa Basin of New Mexico*. Publications in Anthropology 10. El Paso Centennial Museum, University of Texas, El Paso. Historic and Natural Resources Report No. 3. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Caro, T. M., R. Roper, M. Young, and G. R. Dank
1979 Inter-Observer Reliability. *Behaviour* 69(3–4):303–315.

Childs, S. Terry
1995 The Curation Crisis. *Federal Archaeology* 7(4):11–15.

Church, Tim, Mark Ennes, and Mark Sale
2002 *Production and Subsistence in the Shadow of the Jarilla Mountains: Evaluation of 383 Prehistoric Sites in Maneuver Areas 7A and 7B, Fort Bliss, New Mexico*. Report No. 515. Lone Mountain Archaeological Services, El Paso. Historic and Natural Resources Report No. 99-10. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Clark, Jeffery J.
2001 *Tracking Prehistoric Migrations: Pueblo Settlers among the Tonto Basin Hohokam*. Anthropological Papers No. 65. University of Arizona Press, Tucson.

Clark, Jeffery J., Patrick D. Lyons, J. Brett Hill, Henry D. Wallace, William H. Doelle, and Stacey Lengyel
2006 Of Migrants and Mound-Builders. In *Migrants and Mounds: Classic Period Archaeology in the Lower San Pedro Valley*, edited by Jeffery J. Clark, Patrick D. Lyons, and J. Brett Hill. Anthropological Papers No. 45. Center for Desert Archaeology, Tucson.

Cohen, Jacob
1960 A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 220:37–46.

Colton, Harold S., and Lynn L. Hargrave
1937 *Handbook of North American Pottery Wares.* Bulletin No. 11. Museum of Northern Arizona, Flagstaff.

Cook, Patricia
2003 *The Braden Tusk: Discovery and Excavation of a Mammoth Tusk on the East Range of the Fort Huachuca Military Reservation, Cochise County, Arizona*. Project Report No. 03-139. Desert Archaeology, Tucson.

2004 *The Soldier Creek Survey: A Cultural Resources Survey on the East Range, Fort Huachuca Military Reservation, Cochise County, Arizona*. Technical Report No. 2003-04. Desert Archaeology, Tucson.

2005 *An Archaeological Survey of Planning Area 4, Fort Huachuca Military Reservation, Cochise County, Arizona.* Technical Report No. 2004-14. Desert Archaeology, Tucson.

Crown, Patricia L.
1994 *Ceramics and Ideology: Salado Polychrome Pottery.* University of New Mexico Press, Albuquerque.

Daniels, S. G. H.
1972 Research Design Models. In *Models in Archaeology*, edited by David L. Clark, pp. 201–229. Methuen, London, England.

Dean, Jeffrey S.
1991 Thoughts of the Hohokam Chronology. In *Exploring the Hohokam: Prehistoric Desert Peoples of the American Southwest*, edited by George J. Gumerman, pp. 61–149. University of New Mexico Press, Albuquerque.

Dean, Jeffrey S. (editor)
2000 *Salado*. Amerind Foundation New World Studies Series No. 4. University of New Mexico Press, Albuquerque.

Deaver, William L., and Richard S. Ciolek-Torrello
1995 Early Formative Period Chronology for the Tucson Basin. *Kiva* 60:481–529.

Dering, J. Phil
2005 Ecological Factors Affecting the Late Archaic Economy of the Lower Pecos River Region. In *The Late Archaic across the Borderlands: From Foraging to Farming*, edited by Bradley J. Vierra, pp. 247–258. Texas Archaeology and Ethnohistory Series, Thomas R. Hester, general editor. University of Texas Press, Austin.

Dibble, Harold L., and Mary C. Bernard
1980 A Comparative Study of Basic Edge Angle Measurement Techniques. *American Antiquity* 45:857–865.

Diehl, Michael W. (editor)
2005 *Subsistence and Resource Use Strategies of Early Agricultural Communities in Southern Arizona*. Anthropological Papers No. 34. Center for Desert Archaeology, Tucson.

Di Peso, Charles C.
1951 *The Babocomari Village Site on the Babocomari River, Southeastern Arizona*. Publication No. 5. Amerind Foundation, Dragoon, Arizona.

1958 Western Pueblo Intrusion into the San Pedro Valley. *The Kiva* 23:12–16.

Doolittle, William E., and Jonathan B. Mabry
2006 Environmental Mosaics, Agricultural Diversity, and the Evolutionary Adoption of Maize in the American Southwest. In *Histories of Maize: Multidisciplinary Approaches to the Prehistory, Linguistics, Biogeography, Domestication, and Evolution of Maize*, edited by John E. Staller, Robert H. Tykot, and Bruce F. Benz, pp. 109–121. Elsevier Academic, Burlington, Massachusetts.

Doolittle, William E., and James A. Neely
2004 *The Safford Valley Grids: Prehistoric Cultivation in the Southern Arizona Desert*. University of Arizona Press, Tucson.

Faunce, Kenneth V.
2000 *The Perception of Landscape in the Use and Settlement of the Tularosa Basin, New Mexico*. Ph.D. dissertation, College of Graduate Studies, University of Idaho, Moscow. University Microfilms, Ann Arbor, Michigan.

Fish, Paul R.
1978 Consistency in Archaeological Measurement and Classification: A Pilot Study. *American Antiquity* 43:86–89.

Fort Bliss
2012 *Processing and Curation of Collections: Requirements for Submission of Archaeological Collections to the Fort Bliss Curatorial Facility*. Updated January 23, 2012. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Friedman, Gerald M.
1958 Determination of Sieve-Size Distribution from the Thin Section Data for Sedimentary Petrological Studies. *Journal of Geology* 66(4):394–416.

Gavan, James A.
1950 The Consistency of Anthropometric Measurements. *American Journal of Physical Anthropology* 8:417–426.

Gnaden, Denis, and Simon Holdaway
2000 Understanding Observer Variation When Recording Stone Artifacts. *American Antiquity* 65:739–747.

Goldborer, S. Eileen
1985 Report on the Botanical Remains from Conejo Village. In *An Evaluation of a Mesilla Phase Land Use Model on Fort Bliss*, by Robert J. Hard. Draft. On file, Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Graesch, Anthony P.
2009 Fieldworker Experience and Single-Episode Screening as Sources of Data Recovery Bias in Archaeology: A Case Study from the Central Pacific Northwest Coast. *American Antiquity* 74:759–779.

Griset, Suzanne, and Marc Kodack
1999 *Guidelines for the Field Collections of Archaeological Materials and Standard Operating Procedures for Curating Department of Defense Archaeological Collections*. Legacy Project No. 98-1714. Mandatory Center of Expertise of the Curation and Management of Archaeological Collections, U.S. Army Corps of Engineers, St. Louis.

Grissino-Mayer, Henri D., Christopher H. Baisan, and Thomas W. Swetnam
1997 *A 1,373-Year Reconstruction of Annual Precipitation for the Southern Rio Grande Basin*. Department of Physics, Astronomy, and Geosciences, Valdosta State University, Georgia, and Laboratory of Tree-Ring Research, University of Arizona, Tucson. Submitted to the Leg-

acy Program, Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Hanselka, J. Kevin, Bradley J. Vierra, and Kari M. Schmidt (editors)
2010 *Results of an 11,514-Acre Cultural Resource Survey on Northern McGregor Range, Fort Bliss Military Reservation, Otero County, New Mexico*. Technical Report 09-47. Statistical Research, El Paso. Historic and Natural Resources Report No. 08-26. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Hard, Robert J., Pamela Graeber, Jimmie Manasco, Cynthia L. Tennis, and Kevin Thuesen
1994 Doña Ana Phase Ceramics and Cultural Evolution in the Southern Jornada Mogollon. *Bulletin of the Texas Archeological Society* 62:267–283.

Hard, Robert J., Raymond P. Mauldin, and Gerry R. Raymond
1996 Mano Size, Stable Carbon Isotope Ratios, and Macrobotanical Remains as Multiple Lines of Evidence of Maize Dependence in the American Southwest. *Journal of Archaeological Method and Theory* 3(3):253–318.

Hard, Robert J., and John R. Roney
2005 The Transition to Farming on the Rio Casas Grandes and in the Southern Jornada Mogollon Region. In *The Late Archaic across the Borderlands: From Foraging to Farming*, edited by Bradley J. Vierra, pp. 141–186. Texas Archaeology and Ethnohistory Series, Thomas R. Hester, general editor. University of Texas Press, Austin.

Harris, Charles H., and Louis R. Sadler
1993 *Bastion on the Border: Fort Bliss, 1854–1943*. Historic and Natural Resources Report No. 6. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Haury, Emil W.
1976 *The Hohokam: Desert Farmers and Craftsmen, Excavations at Snaketown, 1964–1965*. University of Arizona Press, Tucson.

Haynes, C. Vance, Jr.
1993 Clovis-Folsom Geochronology and Climatic Change. In *From Kostenki to Clovis: Upper Paleolithic–Paleo-Indian Adaptations*, edited by Olga Soffer and N. D. Praslov, pp. 219–236. Plenum, New York.

Heckman, Robert A., Barbara K. Montgomery, and Stephanie M. Whittlesey
2000 *Prehistoric Painted Pottery of Southeastern Arizona*. Technical Series 77. Statistical Research, Tucson.

Heilen, Michael, Christopher L. Nagle, and Jeffrey H. Altschul
2008 *An Assessment of Archaeological Data Quality: A Report Submitted in Partial Fulfillment of Legacy Resource Management Program Project "To Develop Analytical Tools for Characterizing, Visualizing, and Evaluating Archaeological Data Quality Systematically for Communities of Practice within the Department of Defense."* Technical Report 08-65. Statistical Research, Tucson.

Heilen, Michael P., Lynne Sebastian, Jeffrey H. Altschul, Phillip O. Leckman, and Adam Byrd
2012 *Modeling of Archaeological Site Location and Significance at White Sands Missile Range, New Mexico*. Technical Report 12-06. Statistical Research, Tucson.

Heyman, Ira Michael
1997 Smithsonian Perspectives: Using and Taking Care of 140 Million Items. Electronic document, http://siarchives.si.edu/collections/siris_sic_3669, accessed March 5, 2013.

Hill, J. Brett, Jeffery J. Clark, William H. Doelle, and Patrick D. Lyons
2004 Prehistoric Demography in the Southwest: Migration, Coalescence, and Hohokam Population Decline. *American Antiquity* 69:689–716.

Hoffman, Charles Marshall
1997 Alliance Formation and Social Interaction during the Sedentary Period: A Stylistic Analysis of Hohokam Arrowpoints. Unpublished Ph.D. dissertation, Department of Anthropology, Arizona State University, Tempe.

Holliday, Vance T.
1997 *Paleoindian Geoarchaeology of the Southern High Plains*. Texas Archaeology and Ethnohistory Series, Thomas R. Hester, general editor. University of Texas Press, Austin.

Huckell, Bruce B.
1982 *The Distribution of Fluted Points in Arizona: A Review and an Update*. Archaeological Series No. 145. Cultural Resource Management Division, Arizona State Museum, University of Arizona, Tucson.

1984 *The Archaic Occupation of the Rosemont Area, Northern Santa Rita Mountains, Southeastern Arizona*. Archaeological Series No. 147. Cultural Resource Management Division, Arizona State Museum, University of Arizona, Tucson.

1995 *Of Marshes, Maize, and Men: Preceramic Agricultural Settlements in the Cienega Valley, Southeastern Arizona*. Anthropological Papers No. 59. University of Arizona Press, Tucson.

Huckell, Bruce B., and C. Vance Haynes, Jr.
2007 Clovis Paleoecology as Viewed from Murray Springs, Arizona. In *Murray Springs: A Clovis Site with Multiple Activity Areas in the San Pedro Valley, Arizona*, edited by C. Vance Haynes, Jr., and Bruce B. Huckell, pp. 214–225. Anthropological Papers No. 71. University of Arizona Press, Tucson.

Hunter, Rosemary
1988 The Tony Colon Site I. In *Fourth Jornada Mogollon Conference (Oct. 1985) Collected Papers*, edited by Meliha S. Duran and Karl L. Laumbach, pp. 137–162. Human Systems Research, Tularosa, New Mexico.

Jamieson, Perry
1993 *A Survey History of Fort Bliss, 1890–1940*. Historic and Natural Resources Report No. 5. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Jamison, Paul L., and Stephen L. Zegura
1974 A Univariate and Multivariate Examination of Measurement Error in Anthropometry. *American Journal of Physical Anthropology* 40(2):197–203.

Kelley, Jane H.
1984 *The Archaeology of the Sierra Blanca Region of Southeastern New Mexico*. Anthropological Papers No. 74. Museum of Anthropology, University of Michigan, Ann Arbor.

Kemrer, Meade F.
2006 Puebloan Integration: An Example from South Central New Mexico. Paper presented at the Mogollon Conference, Tucson.

Kemrer, Meade F. (editor)
2008 *Agriculture in the Southern San Andres Mountains, A.D. 900–1400, South Central New Mexico*. Environmental Document Report No. 552. Environmental Stewardship Division Volunteers, White Sands Missile Range, New Mexico.

Kludt, Trevor J.
2007 *Landform and Settlement, Part I: Farming the Fans*. Report No. 560-016-017. Lone Mountain Archaeological Services, El Paso.

Lentz, Stephen C.
2006 *High Rolls Cave: Insectos, Burritos, y Frajos: Archaic Subsistence in Southern New Mexico: Excavations at LA 114103, Otero County, New Mexico*. Archaeology Notes 345. Office of Archeological Studies, Museum of New Mexico, Santa Fe.

Lindsay, Alexander J., Jr.
1987 Anasazi Population Movements to Southeastern Arizona. *American Archaeology* 6(3):190–198.

Liritzis, Ioannis, and Maria Diakostamatiou
2002 Towards a New Method of Obsidian Hydration Dating with Secondary Ion Mass Spectrometry via a Surface Saturation Layer Approach. *Mediterranean Archaeology and Archaeometry* 2(1):3–20.

Liritzis, Ioannis, and Christopher M. Stevenson (editors)
2012 *Obsidian and Ancient Manufactured Glasses*. University of New Mexico Press, Albuquerque.

Little, Barbara J., Erika Martin Seibert, Jan Townsend, John H. Sprinkle, Jr., and John Knoerl
2000 *Guidelines for Evaluating and Registering Archaeological Properties*. Archaeological Assistance Division, U.S. Department of the Interior National Park Service, Cultural Resources, Washington, D.C.

Lyman, R. Lee, and Todd L. VanPool
2009 Metric Data in Archaeology: A Study of Intra-Analyst and Inter-Analyst Variation. *American Antiquity* 74:485–504.

Lyons, Patrick D.
2003 *Ancestral Hopi Migrations*. Anthropological Papers No. 68. University of Arizona Press, Tucson.

Lyons, Patrick D., E. Charles Adams, Jeffrey H. Altschul, C. Michael Barton, and Chris M. Roll
2006 *The Archaeological Curation Crisis in Arizona: Analysis and Possible Solutions.* A Report Prepared by the Governor's Archaeology Advisory Commission Curation Subcommittee. Electronic document, http://azstateparks.com/committees/downloads/GAAC_Curation_Crisis_Full.pdf, accessed March 5, 2013.

Mabry, Jonathan B.

1998 *Paleoindian and Archaic Sites in Arizona*. Technical Report No. 97-7. Center for Desert Archaeology, Tucson.

2005 Changing Knowledge and Ideas about the First Farmers in Southeastern Arizona. In *The Late Archaic across the Borderlands: From Foraging to Farming*, edited by Bradley J. Vierra, pp. 41–83. Texas Archaeology and Ethnohistory Series, Thomas R. Hester, general editor. University of Texas Press, Austin.

MacNeish, Richard S. (editor)

1993 *Preliminary Investigations of the Archaic in the Region of Las Cruces, New Mexico*. Historic and Natural Resources Report No. 9. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

MacWilliams, A. C., Bradley J. Vierra, and Kari M. Schmidt (editors)

2009 *Archaeological Mitigation at FB 17 (LA 91017) and FB 9122 (LA 30116) on the Doña Ana Range, Fort Bliss, Doña Ana County, New Mexico*. Technical Report 09-13. Statistical Research, El Paso. Historic and Natural Resources Report No. 07-49. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Majewski, Teresita, and Michael J. O'Brien

1987 The Use and Misuse of Nineteenth-Century English and American Ceramics in Archaeological Analysis. *Advances in Archaeological Method and Theory* 11:97–209.

Mauldin, Raymond P.

1986 Settlement and Subsistence Patterns during the Pueblo Period on Fort Bliss, Texas: A Model. In *Mogollon Variability*, edited by Charlotte Benson and Steadman Upham, pp. 255–270. University Museum Occasional Papers No. 15. New Mexico State University, Las Cruces.

McMaster, Richard K.

1962 *Musket, Saber, and Missile: A History of Fort Bliss*. R. K. McMaster, El Paso.

Metz, Leon Claire

1988 *Desert Army: Fort Bliss on the Texas Border*. Mangan Books, El Paso.

Miller, Myles R.

2005 Architecture and Social Integration in Jornada Mogollon Pueblos. Paper presented at the 14th Biennial Jornada Mogollon Conference, El Paso Museum of Archaeology, El Paso.

Miller, Myles R., and Chad Burt (editors)

2007 *Miscellaneous Investigations at the Conejo Site (LA 91044/FB 46), Fort Bliss, Doña Ana County, New Mexico*. Report of Investigations No. 680B-EP. Geo-Marine, El Paso. Historic and Natural Resources Report No. 03-17b. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Miller, Myles R., and Timothy B. Graves

2009 *Madera Quemada Pueblo: Archaeological Investigations of a Fourteenth Century Jornada Mogollon Pueblo.* Report of Investigations No. 679-EP. Geo-Marine, El Paso. Historic and Natural Resources Report No. 03-12. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Miller, Myles R., Timothy B. Graves, Moira Ernst, and Michael Stowe
2011 *Burned Rock Middens of the Southern Sacramento Mountains*. Report of Investigations No. 782-EP. Geo-Marine, El Paso. Historic and Natural Resources Report No. 09-28. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Miller, Myles R., and Nancy A. Kenmotsu
2004 Prehistory of the Jornada Mogollon and Eastern Trans-Pecos Regions of West Texas. In *The Prehistory of Texas*, edited by Timothy K. Perttula, pp. 205–265. Anthropology Series No. 9. Texas A&M University Press, College Station.

Miller, Myles R., Nancy A. Kenmotsu, and Melinda R. Landreth (editors)
2009 *Significance and Research Standards for Prehistoric Archaeological Sites at Fort Bliss: A Design for the Evaluation, Management, and Treatment of Cultural Resources*. Report of Investigations No. 697-EP. Geo-Marine, El Paso. Historic and Natural Resources Report No. 05-16. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Mizutani, Shinjiro
1963 A Theoretical and Experimental Consideration on the Accuracy of Sieving Analysis. *Journal of Earth Sciences* 11:1–27.

Nelson, Margaret C.
1999 *Mimbres during the Twelfth Century: Abandonment, Continuity, and Reorganization*. University of Arizona Press, Tucson.

Nepstad-Thornberry, Tina, Curtis Nepstad-Thornberry, Melissa Stoltz, Karyn de Dufour, and Richard Wilshusen
2002 Addressing the Curation Crisis in Colorado: An Assessment for the Executive Committee of the Colorado Council of Professional Archaeologists. University of Colorado Museum, Boulder. Electronic document, http://www.sha.org/documents/research/collections_management/Nepsted-thornberryAddressingtheCurationCrisisinColorado.pdf, accessed March 5, 2013.

Neuman, J., J. W. Chardine, and J. M. Porter
1999 Approaches to Testing Inter-Observer Reliability of Field-Collected Behavioral Data. *Waterbirds: The International Journal of Waterbird Biology* 22(3):348–357.

New Mexico Wilderness Alliance
2013 Otero Mesa: America's Wildest Grassland. Electronic document, http://www.oteromesa.org/the-campaign-to-protect-otero-mesa/, accessed February 27, 2013.

Opler, Morris E., and Catherine H. Opler
1950 Mescalero Apache History in the Southwest. *New Mexico Historical Review* 25:1–36.

Parker, Patricia L., and Thomas F. King
1998 *Guidelines for Evaluating and Documenting Traditional Cultural Properties*. National Register Bulletin 38. Rev. ed. Interagency Resources Division, U.S. Department of the Interior National Park Service, Washington, D.C. Available online at http://www.nps.gov/nr/publications/bulletins/nrb38/, accessed March 5, 2013.

Railey, Jim A. (editor)
2010 *Archaeological Data Comparability for the Permian Basin Mitigation Program*. Permian Basin Mitigation Program, Task 4. Project No. 16362. SWCA Environmental Consultants, Albuquerque. Prepared for the U.S. Department of the Interior Bureau of Land Management, Carlsbad Field Office, Carlsbad, New Mexico, Contract No. L08PC90396, Legacy No. NAC080148, Delivery Order No. L10PD02650.

Ravesloot, John C., and Stephanie M. Whittlesey
1987 Inferring the Protohistoric Period in Southern Arizona. In *The Archaeology of the San Xavier Bridge Site (AZ BB:13:14) Tucson Basin, Southern Arizona*, edited by John C. Ravesloot, pp. 81–98. Archaeology Series 171. Arizona State Museum, University of Arizona, Tucson.

Reff, Daniel T.
1991 *Disease, Depopulation, and Culture Change in Northwestern New Spain, 1518–1764*. University of Utah Press, Salt Lake City.

Riciputi, Lee R., J. Michael Elam, Lawrence M. Anovitz, and David R. Cole
2002 Obsidian Diffusion Dating by Secondary Ion Mass Spectrometry: A Test Using Results from Mound-65, Chalco, Mexico. *Journal of Archaeological Science* 29(10):1055–1075.

Rolak, Bruno
1974 History of Fort Huachuca, 1877–1890: The Apache Years. *The Smoke Signal* 29:206–224.

Roney, John R., and Robert J. Hard
2002 Early Agriculture in Northwestern Chihuahua. In *Traditions, Transitions, and Technologies: Themes in Southwestern Archaeology*, edited by Sarah Schlanger, pp. 160–177. University of Colorado Press, Boulder.

Sagebiel, Kerry L., Kelly L. Jenks, Teresita Majewski, and Lauren E. Jelinek
2007 *Archaeological Collections Management Procedures: Department of Defense Legacy Resource Management Program.* Technical Report 07-41. Draft. Statistical Research, Tucson. U.S. Army Corps of Engineers, Los Angeles District, Los Angeles, and the U.S. Army Engineer Research and Development Center, Construction Engineering Research Laboratory, Vicksburg, Mississippi.

Shrimpton, Rebecca H. (editor)
2002 *How to Apply the National Register Criteria for Evaluation*. Revised for the Internet and finalized by Patrick W. Andrus. Electronic document, http://www.nps.gov/nr/publications/bulletins/nrb15/, accessed January 9, 2013.

Society for American Archaeology (SAA) Advisory Committee on Curation
2003 The Archaeological Curation Crisis: An Integrated Action Plan for the SAA and Its Partners. Available online at http://pdfdownloadfree.net/?pdfurl=1qeXpurpn6Wih-SUpOGunK6nh8DX24W12sjY087d4NXc1sjN25ao6drG5NvY3JSp59bY1eKwhbXWhbng3dPb2NbhytCPt8jo0dTekrna1dSV09Tej6STooqg46yaqJbMl6jYpqOrn5bQ2eTio52j2enOy9Kd7cbn0M7e2d3d4pTa0dqb4trK4tzU3qHcz9Xd5ZvV0NWYoPE, accessed March 5, 2013.

Sayles, Edwin B.
1945 *Material Culture: The San Simon Branch Excavations at Cave Creek and in the San Simon Valley, Vol. 1*. Medallion Papers 34. Gila Pueblo, Globe, Arizona.

Sayles, Edwin B., and Ernst Antevs

1941 *The Cochise Culture*. Medallion Papers 39. Gila Pueblo, Globe, Arizona.

Sebastian, Lynne

2009 Deciding What Matters: Archaeology, Eligibility, and Significance. In *Archaeological and Cultural Resource Management: Visions for the Future*, edited by Lynne Sebastian and William D. Lipe, pp. 91–114. School for Advanced Research Press, Santa Fe.

Seymour, Deni J.

1989 The Dynamics of Sobaipuri Settlement in the Eastern Pimeria Alta. *Journal of the Southwest* 31:205–222.

2002 *Conquest and Concealment: After the El Paso Phase on Fort Bliss: An Archaeological Study of the Manso, Suma, and Early Apache*. Report Nos. 525 and 528. Lone Mountain Archaeological Services, El Paso. Historic and Natural Resources Report No. 01-06. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Shackley, M. Stephen

2005 *Obsidian: Geology and Archaeology in the North American Southwest*. University of Arizona Press, Tucson.

Shoukri, Mohamed M.

2010 *Measures of Interobserver Agreement and Reliability*. 2nd ed. Biostatistics Series. Chapman & Hall/CRC Press, Boca Raton, Florida.

Sim, Julius, and Chris C. Wright

2005 The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy* 85(3):257–268.

Smith, Steven D.

2001 *The African American Soldier at Fort Huachuca, 1892–1946*. South Carolina Institute of Archaeology and Anthropology, University of South Carolina, Columbia.

Spielmann, Katherine (editor)

1998 *Migration and Community Reorganization: The Pueblo IV Period in the American Southwest*. Anthropological Research Papers No. 51. Arizona State University, Tempe.

Stanford, Dennis J.

2005 Paleoindian Archaeology and Late Pleistocene Environments in the Plains and Southwestern United States. In *Ice Age Peoples of North America: Environments, Origins, and Adaptations of the First Americans*, edited by Robson Bonnichsen and Karen L. Turnmire, pp. 281–339. Center for the Study of the First Americans, Department of Anthropology, Texas A&M University, College Station.

Sullivan, Lynne P.

1992 *Managing Archeological Resources from the Museum Perspective*. Technical Brief No. 13. Archaeological Assistance Division, U.S. Department of the Interior National Park Service, Washington, D.C.

Swarthout, Jeanne K., and Alan Dulaney
1982 *A Description of Ceramic Collections from the Railroad and Transmission Line Corridors*. Research Paper 26. Museum of Northern Arizona, Flagstaff.

Tagg, Martyn D.
1996 Early Cultigens from Fresnal Shelter, Southeastern New Mexico. *American Antiquity* 61:311–324.

2011 *1 Sept 2011, East Range Training Area Remastication, Cultural Resources Assessment*. Cultural Resources Report FH-11-9. U.S. Army Intelligence Center, Fort Huachuca, Arizona.

Thiel, J. Homer
2005 *An Assessment of Five Buildings at Fort Huachuca, Cochise County, Arizona*. Project Report No. 05-114. Desert Archaeology, Tucson.

Trimble, Michael K., and Thomas B. Meyers
1991 *Saving the Past from the Future: Archaeological Curation in the St. Louis District.* U.S. Army Corps of Engineers, St. Louis District, St. Louis. Available online at http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA323819, accessed March 5, 2013.

Trousil, Jeff
2001 *Integrated Natural Resources Management Plan and Environmental Assessment.* U.S. Army Intelligence Center, Fort Huachuca, Arizona, and Universe Technologies, Frederick, Maryland.

Upham, Steadman, Richard S. MacNeish, Walton C. Galinat, and Christopher M. Stevenson
1987 Evidence Concerning the Origin of Maize de Ocho. *American Anthropologist* 89:410–419.

U.S. Army (Army)
2007 *Environmental Quality: Environmental Protection and Enhancement*. Army Regulation 200–1. Department of the Army Headquarters, Washington, D.C.

U.S. Department of the Interior National Park Service (NPS)
2011 Archaeology Program. Archeology for Interpreters: Glossary. Electronic document, http://www.nps.gov/archeology/afori/glossary.htm, accessed September 9, 2012.

U.S. Fish and Wildlife Service
2007 Draft Biological Opinion of June 14, 2007, from U.S. Fish and Wildlife Service, Phoenix Arizona, to Colonel Jonathan B. Hunter, Commander, U.S. Army Garrison, Fort Huachuca, Arizona. Reference AESO/SE, 22410-2007-F-0132, 02-21-02-F-229, 02-21-98-F-266. Available online at http://www.fws.gov/southwest/es/arizona/Documents/Biol_Opin/070132_FortHuachucaFinal.pdf , accessed March 10, 2008.

Vanderpot, Rein
1994 *A 6,800-Acre Intensive Survey of Proposed FTX and Other Training Areas on Fort Huachuca, Arizona.* Technical Report 93-20. Statistical Research, Tucson.

1997 From Foraging to Farming: Prehistoric Settlement and Subsistence Dynamics in the San Pedro Valley. In *Prehistory of the Borderlands: Recent Research in the Archaeology of Northern Mexico and the Southern Southwest*, edited by John Carpenter and Guadalupe Sanchez, pp. 33–45. Archaeological Series No. 186. Arizona State Museum, University of Arizona, Tucson.

2012 Project Contexts. In *A 4,022-Acre Intensive Survey, and NRHP-Eligibility Evaluations of 27 Previously Recorded Sites on Fort Huachuca, Arizona*, edited by Rein Vanderpot and William M. Graves, pp. 2.1–2.29. Technical Report 12-31. Draft. Statistical Research, Tucson. Cultural Resources Report FH-12-07. U.S. Army Intelligence Center, Fort Huachuca, Arizona.

Van West, Carla R., Mark T. Swanson, and Jeffrey H. Altschul

1997 *Cultural Resources Management Plan for Fort Huachuca Military Reservation, Arizona*. Technical Series 67. Statistical Research, Tucson.

Vierra, Bradley J.

2007 Archaic Foragers of the Northern Rio Grande Valley, New Mexico. Paper presented at the 72nd Annual Meeting of the Society for American Archaeology, Austin.

2012 Archaeological Background. In *Mitigation of Three Archaeological Sites along and near El Paso Draw in the New IBCT Training Area, East McGregor Range, Fort Bliss Military Reservation, Otero County, New Mexico*, edited by Bradley J. Vierra and Christine G. Ward, pp. 27–43. Technical Report 12-56. Statistical Research, El Paso. Historic and Natural Resources Report No. 11-14. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Vierra, Bradley J., J. Kevin Hanselka, and Jason D. Windingstad

2010 Research Questions. In *Results of an 11,514-Acre Cultural Resource Survey on Northern McGregor Range, Fort Bliss Military Reservation, Otero County, New Mexico*, edited by J. Kevin Hanselka, Bradley J. Vierra, and Kari M. Schmidt, pp. 401–410. Technical Report 09-47. Statistical Research, El Paso. Historic and Natural Resources Report No. 08-26. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Vierra, Bradley J., Phillip O. Leckman, and Robert A. Heckman

2010 *Plan of Work for the Mitigation of Three Archaeological Sites in the New IBCT Training Area, East McGregor Range, Otero County, New Mexico*. Technical Report 10-74. Statistical Research, El Paso. Submitted to the Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

2011 *Plan of Work for the Mitigation of Three Archaeological Sites in the New IBCT Training Area, East McGregor Range, Otero County, New Mexico*. Technical Report 11-31. Statistical Research, El Paso. Submitted to the Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Vierra, Bradley J., and Christine G. Ward

2012 *Mitigation of Three Archaeological Sites along and near El Paso Draw in the New IBCT Training Area, East McGregor Range, Fort Bliss Military Reservation, Otero County, New Mexico*. Technical Report 12-56. Statistical Research, El Paso. Historic and Natural Resources Report No. 11-14. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Wegener, Robert M., Scott O'Mack, Heather Miljour, Scott Thompson, and Richard S. Ciolek-Torrello

2005 *Archaeological Investigations along U.S. 60: Preliminary Results of Phase 2 Data Recovery at Eight Sites along U.S. 60, Queen Valley to Queen Creek Bridge, Pinal County, Arizona*. Technical Report 06-43. Statistical Research, Tucson.

Whalen, Michael E.
1977 *Settlement Patterns of the Eastern Hueco Bolson.* Publications in Anthropology No. 4. El Paso Centennial Museum, University of Texas, El Paso. Historic and Natural Resources Report No. 76-01. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

1978 *Settlement Patterns of the Western Hueco Bolson.* Publications in Anthropology No. 6. El Paso Centennial Museum, University of Texas, El Paso. Historic and Natural Resources Report No. 76-02. Directorate of Public Works, Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas. Submitted to the U.S. Army Corps of Engineers, Forth Worth District, Fort Worth.

Whittaker, John C., Douglas Caulkins, and Kathryn A. Kamp
1998 Evaluating Consistency in Typology and Classification. *Journal of Archaeological Method and Theory* 5(2):129–164.

Whittlesey, Stephanie M., Richard S. Ciolek-Torrello, and Jeffrey H. Altschul (editors)
1997 *Overview, Synthesis, and Conclusions*. Vanishing River: Landscapes and Lives of the Lower Verde Valley: The Lower Verde Archaeological Project. SRI Press, Tucson.

Whittlesey, Stephanie M., Richard S. Ciolek-Torrello, and Matthew A. Sterner
1994 *Southern Arizona the Last 12,000 Years: A Cultural-Historic Overview for the Western Army National Guard Aviation Training Site*. Technical Series 48. Statistical Research, Tucson.

Whittlesey, Stephanie M., and Robert A. Heckman
2000 Culture History and Research Background. In *Prehistoric Painted Pottery in Southeastern Arizona*, by Robert A. Heckman, Barbara K. Montgomery, and Stephanie M. Whittlesey, pp. 1–22. Technical Series 77. Statistical Research, Tucson.

Wiseman, Regge N.
1997 A Preliminary Look at Evidence for Late Prehistoric Conflict in Southeastern New Mexico. In *Layers of Time: Papers in Honor of Robert H. Weber*, edited by Meliha S. Duran and David T. Kirkpatrick, pp. 135–146. Report No. 23. Archaeological Society of New Mexico, Albuquerque.

Woodson, M. Kyle
1999 Migrations in Late Anasazi Prehistory: The Evidence from the Goat Hill Site. *Kiva* 65:63–84.

Worcester, Donald E.
1941 Early Spanish Accounts of the Apache Indians. *American Anthropologist* 43:308–312.

1979 *The Apaches: Eagles of the Southwest.* University of Oklahoma Press, Norman.

APPENDIX A

Research Design for the Surface Collection of Artifacts at Fort Huachuca, in Support of a Legacy Resource Management Program Project to Assess the Quality of In-Field Artifact Analysis

October 20, 2011

Project Context and Goals

Driven by the curation crisis that has affected all archaeological research in the United States, there has been a recent trend among some western states to limit field collections (Lyons et al. 2006; Sagebiel et al. 2010). As a consequence, in-field artifact analysis is increasingly used as an alternative to artifact collection, laboratory analysis, and subsequent curation. But how accurate are in-field analyses? What classes of artifacts can be recorded adequately in the field? Which classes need laboratory analysis? And which artifact classes are likely to be revisited by future researchers in the laboratory? The answer to these questions is, “No one knows.”

Statistical Research, Inc. (SRI), recently received funding from the Department of Defense (DoD) Legacy Resource Management Program to conduct a pilot study of in-field artifact analysis (Project No. 07-353). The purpose of this project is to provide the DoD with good answers to the above questions from which decisions about field recording and future curation can be made. The objective of this project is twofold: (1) to perform a pilot test of the accuracy and adequacy of in-field artifact analysis using surface materials from sites at two U.S. Army installations (Fort Bliss and Fort Huachuca) and (2) to prepare a report of the results that will include a set of best practices for installations and ranges to help guide cultural resource managers (CRM) in the western United States. The ultimate goal of the project is to present the circumstances and the material classes of artifacts that are best suited for in-field analysis. Installation managers will then be in a position to make sound judgments on inventory, evaluation, and data recovery strategies that meet the DoD’s compliance objectives and stewardship responsibilities.

Project Methods

A sample site or sites considered to be likely candidates for in-field analysis by the CRM will be selected at each installation. The selected sites will be visited by multiple teams of archaeologists, who will count the number of surface artifacts, place the artifacts into material classes, type them, and document them with digital photographs. The artifacts will then be collected and analyzed in the laboratory by trained analysts. Additionally, two sets of analysts will independently analyze the collections solely from digital photographs. For each site, the results of the in-field analyses will then be assessed against each other to determine the levels of consistency. The same will be done for the analyses using only digital images. Finally, the in-field analyses and “digital” analyses will be assessed against the laboratory results. Existing reports may also be used to compare the results of initial inventory field assessments of artifacts to subsequent laboratory analysis of artifacts from the same site.

At Fort Huachuca, the Soldier Creek site (AZ EE:7:164 [ASM]) has been identified as an excellent candidate for this project (Cook 2003; Vanderpot 1994:86–92) (Figure A.1). The Soldier Creek Site is a large, intensively utilized Formative period habitation site located near the end of a broad ridge that overlooks a wide, grass-covered floodplain of Soldier Creek. The site was originally recorded by Vanderpot (1994) during survey and was rerecorded during a subsequent survey by Cook (2002). Sediments at the site consist of mixed gravel and loamy alluvium. Vegetation is dominated by mesquite; desert broom, burroweed, and cholla are also present. The site consists of an extensive ceramic and lithic scatter, at least four trash mounds, and two rock rings. An isolated historical-period trash scatter and modern artifacts associated with military activity are also present.

Lithics at the site consist mostly of debitage and cores made on a variety of raw materials, including chert, jasper, hornfels, quartzite, quartz, chalcedony, rhyolite, and andesite. One piece of obsidian was also observed. Flaked stone tools identified at the site included modified flakes, a scraper, core/chopper tools, tool blanks, and a projectile point. Ground stone artifacts—all of which were made on granite, quartzite, or an unidentified material—consisted of at least six fragments from basin metates, a trough-metate fragment,

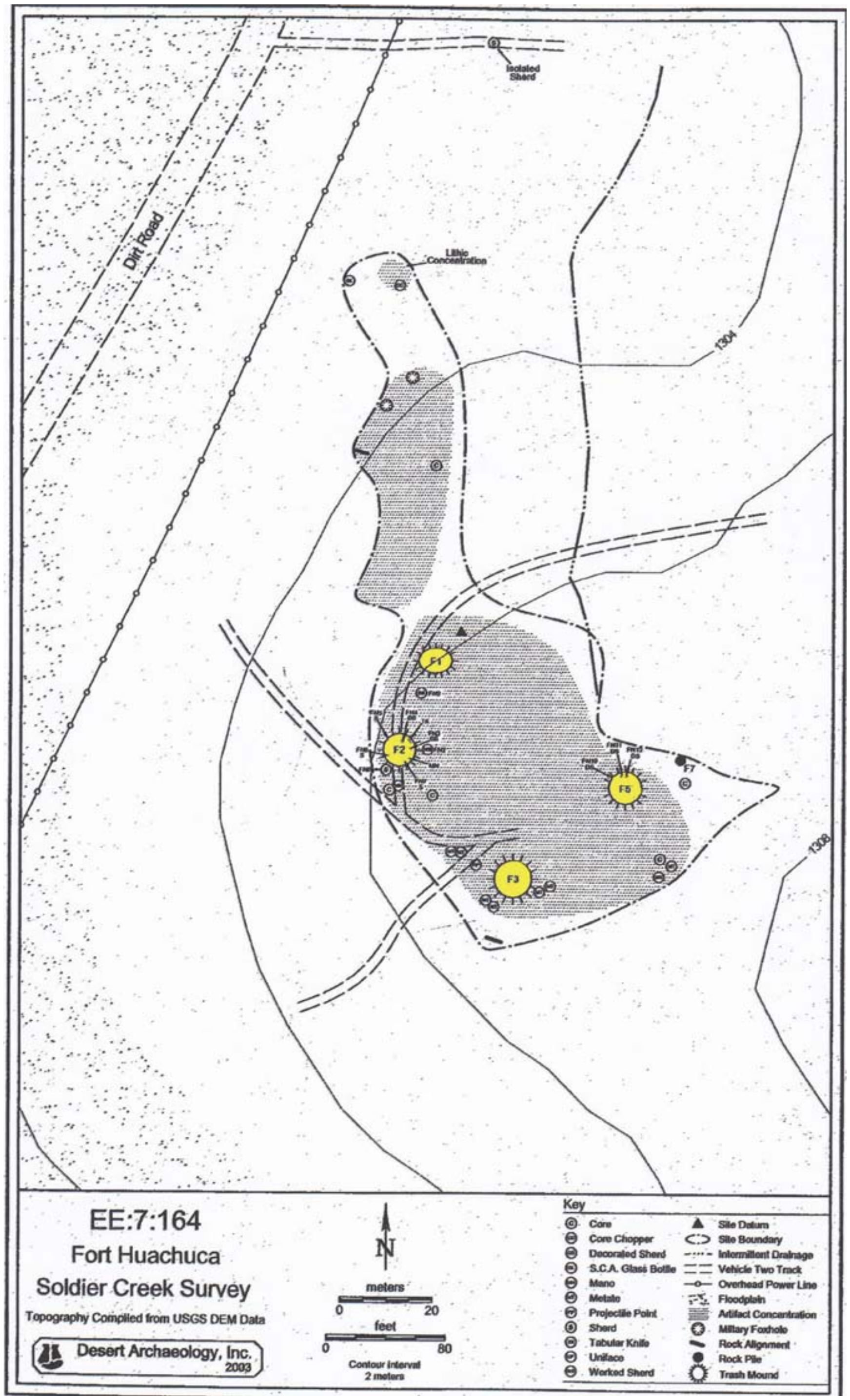

Figure A.1. The Soldier Creek site, showing in yellow highlight the four trash mound features to be surface collected (reproduced from Cook [2002:53, Figure 3.24.])

and four manos. Ceramics consisted of a variety of painted and plain wares, including wares associated with the Tucson Basin Hohokam, Dragoon, and Babocomari ceramic traditions (Heckman et al. 2000). Pit-house features are suspected to be present because of the presence of trash mounds but have not been observed on the site surface.

When first recorded by Vanderpot (1994), thousands of artifacts were noted at the site, including hundreds on the surfaces of individual trash mounds. When rerecorded by Cook (2002), however, the surface characteristics of the site had changed; only 110 flaked stone artifacts, 102 ceramic artifacts, and 7 ground stone artifacts were noted. Erosion was inferred to have caused changes to the site surface between recording episodes. In addition to erosion, the site has been impacted by vehicle traffic and modern military activity, and at least one feature (Feature 2) has been directly impacted by modern activities. Multiple collections have also been made from the site, including a substantial number of decorated sherds.

Surface-artifact counts remained comparatively low during a recent visit to the site by the Fort Huachuca CRM; fewer than 100 artifacts were observed on the surface of any of the four trash mound features. Visibility was excellent. Recent disturbances to the site appeared relatively minor and are due to natural sheet erosion and bioturbation.

Field Methods

Five trash mounds have been previously identified at the Soldier Creek site (Vanderpot 1994), but one has since eroded or is completely obscured by vegetation. Each of the four remaining trash mounds identified at the site (Features 1, 2, 3, and 5) will be sampled with a single collection unit, for a total of four collection units. At the present time, we anticipate that collection units 10 by 10 m in size will encompass the entirety of each trash mound and will be sufficient to collect the artifacts needed for analysis. In all likelihood, the surface collection will result in the collection of fewer than 400 artifacts, a number that should be sufficient to meet the goals of the project.

The Advisory Council on Historic Preservation (ACHP) and the Arizona State Historic Preservation Office (SHPO) agreed that there would be no adverse effect on AZ EE:7:164 (ASM) if artifacts collected during the course of the project were returned to the field by the end of the project, rather than curated. The collected artifacts will be analyzed and photographed in the field and subsequently analyzed in the laboratory before being returned to their original collection units. To ensure that collected artifacts are returned to the site in close proximity to their original positions in the field, each collection unit will be divided evenly into uniquely provenienced, 2-by-2-m grid cells. Photographs that each include a metric scale and a north arrow will be taken of each grid cell in the field prior to artifact collection. The grid cells will serve as the provenience designations for collected artifacts, and artifacts will be returned to these provenience units, using the field photographs as guides to artifact placement.

A mapping-grade global positioning system (GPS) unit loaded with geospatial data representing the locations of previously identified site boundaries, datums, and features will be used at the inception of fieldwork. Once the features and attributes mentioned in previous reports are relocated, the GPS unit will be used to establish or reestablish Universal Transverse Mercator World Geodetic System 1984 coordinates for the site datum and the southwest corner of each collection unit. SRI will allow the GPS unit to record continuously for a minimum of 10 minutes for each measured location, a process that will produce coordinates accurate within 10–15 cm once the GPS data are differentially corrected.

A total of four sample units for surface collection will be placed so as to encompass each of the four trash mounds identified at the site (see Figure 1); each sample unit will measure no more than 7 by 7 m in size. The field crew will place a pin flag adjacent to each artifact that is to be collected, photograph the artifact, and identify the artifact according to type and relevant technological attributes, by material class. Field recording will be done using SRI's system for recording provenience information for artifacts, features, and sample units. Following widely accepted conventions for the locations of artifacts, samples, and activity areas, this system provides a separate list of provenience numbers for each archaeological site and assigns a unique number to every space that contains features, artifacts, or other archaeological units. Digital photographs will include scales and will be taken of the obverse and reverse surfaces of each artifact (e.g., the interior and exterior surfaces of a ceramic artifact) and will be documented such that each individual artifact can be clearly identified and related to its analysis data. At a separate time closely following this initial recording episode (i.e., on the same day or the following day), a different field crew will identify each artifact, using the same recording system and also using the pin flags and digital photographs, if necessary, to relocate each artifact. After being analyzed a second time in the field, the artifacts will be collected following standardized procedures for the collection and storage of artifacts (in accordance with Title 36, Part 79, of the *Code of Federal Regulations* [36 CFR 79] and established curatorial standards and guidelines). Because collections of decorated ceramics and other artifact classes have previously been made from the site, these artifacts will also be made available in the field by the Fort Huachuca CRM for the field crew to analyze and photograph.

During the course of surface collection, there is the possibility that human remains will be encountered. Fort Huachuca will be immediately notified should human remains be encountered during the course of this project. They will be treated with the utmost respect, according to the protocols outlined by the Native American Graves Protection and Repatriation Act (NAGPRA) and addressed in Fort Huachuca's Integrated Cultural Resources Management Plan. There is no plan, however, to collect and analyze human remains as part of this project.

Laboratory Methods

Laboratory processing will follow established SRI guidelines and will meet all the requirements of the Arizona State Museum (ASM). Archival-quality storage materials will be used. Computerized inventories of artifacts, proveniences, and catalog lists will be maintained by the laboratory. Computer files, databases, and inventories will be regularly backed up, and copies of all records will be kept in fireproof facilities. All of the laboratory analyses and recording of artifacts and materials collected will mirror the in-field recording. All of the field and laboratory observations collected will be entered into SRI's relational database, which will provide relational links between individual artifacts and their specific proveniences as well as permit comparisons of identifications made in the field, identifications made using digital photographs, and identifications made from the collected artifacts in the laboratory.

The physical artifacts as well as digital photographs of the artifacts will be analyzed in the laboratory by multiple teams of analysts. If an individual analyst needs to participate in both the analysis of digital photographs and laboratory analysis of physical-artifact specimens, the analyst will first analyze digital photographs of artifacts and then analyze the physical artifacts during a separate recording episode. To ensure independence between the two analyses in such a case, the analyst will not be provided the opportunity to revise identifications made during the digital-photograph analysis after having seen the physical specimens. All ceramic materials collected during this project will be classified into traditional ware and type categories to the finest levels possible. As data permit, the dimensions and shapes of individual vessels will be inferred. Lithic analyses will focus on the identification of artifact types, raw-material types, and technological attributes.

Analysis and Reporting of Project Results

Statistical measures will be employed to assess the levels of consistency among in-field analysis, digital analysis, and laboratory analysis. Assessments will determine whether artifact identifications among the three methods agree or disagree, how often they agree or disagree, and the degree to which assessments vary in the level of precision. For instance, in-field analysis might indicate that a sherd dates to the Sedentary period of the Hohokam sequence (ca. A.D. 900–1150), and a subsequent laboratory analysis might conclude that the same sherd could be more precisely dated to the Rillito phase (ca. A.D. 900–1000) of the Hohokam sequence. Both identifications could be technically accurate, but one is more precise, in that it provides a greater level of specificity.

Ultimately, information derived from artifacts is essential to interpreting site function and temporal and cultural affiliation. In turn, these data are essential to making National Register of Historic Places–eligibility recommendations and decisions regarding level of effort, sampling requirements, and avoidance. The adequacy of in-field analysis will be assessed by determining whether enough accurate information was derived to result in an appropriate identification of the resource. The results of these assessments will be used to identify the kinds of artifacts that are best suited or poorly suited to in-field analysis and to recommend best practices for in-field analysis. A final report of the results will be prepared that will include a set of best practices for western installations and ranges.

Final Disposition of Collected Artifacts and Project Materials

As stated above, the ACHP and the Arizona SHPO agreed that there would be no adverse effect on the Soldier Creek site if the artifacts collected during the course of the project were returned to the field, rather than curated. Following this recommendation, the Fort Huachuca CRM will ensure that all artifacts collected during the course of the project are returned to the provenience units from which they were originally collected.

The Fort Huachuca CRM will also ensure that artifacts previously collected from the site and used during the course of the project for making artifact identifications, along with any records resulting from the project, are curated by the ASM in accordance with standards and guidelines generated by 36 CFR 79, with consideration of any claims or conditions recognized as a result of consultation with Native American tribes according to the provisions of the NAGPRA. Upon receipt of the final report for the project, the Native American tribes shall be provided the opportunity to review any and all recovered materials. Any human remains, associated funerary objects, sacred objects, and objects of cultural patrimony that may have been discovered during work at the Soldier Creek site, including those to be returned, reburied, or otherwise repatriated to those Native American tribes having established claims of affiliation or descent under NAGPRA, shall be treated with dignity and respect at all times. Further, any specific treatments identified by the affiliated Native American tribes for application to claimed materials prior to disposition will be observed.

References

Cook, Patricia

2003 *The Soldier Creek Survey: A Cultural Resources Survey on the East Range, Fort Huachuca Military Reservation, Cochise County, Arizona*. Technical Report 2003-4. Center for Desert Archaeology, Tucson.

Heckman, Robert A., Barbara K. Montgomery, and Stephanie M. Whittlesey

2000 *Prehistoric Painted Pottery of Southeastern Arizona*. Technical Series 77. Statistical Research, Tucson.

Lyons, Patrick D., E. Charles Adams, Jeffrey H. Altschul, C. Michael Barton, and Chris M. Roll

2006 *The Archaeological Curation Crisis in Arizona: Analysis and Possible Solutions*. A Report Prepared by the Arizona Governor's Archaeology Advisory Commission Curation Subcommittee. On file, Arizona State Historic Preservation Office, Phoenix.

Sagebiel, Kerry L., Kelly L. Jenks, Teresita Majewski, and Lauren E. Jelinek

2010 *Archaeological Collections Management Procedures, Legacy Resource Management Program, Project No. 06-319.* Technical Report 07-41. Statistical Research, Tucson.

Vanderpot, Rein

1994 *A 6,800-Acre Intensive Survey of Proposed FTX and Other Training Areas on Fort Huachuca, Arizona*. Technical Report 93-20. Statistical Research, Tucson.