The application and misapplication of mass analysis in lithic debitage studies

William Andrefsky Jr.

Department of Anthropology, Washington State University, Pullman, WA 99164-4910, United States

Received 4 March 2006; received in revised form 25 May 2006; accepted 26 May 2006

Abstract

The technique of debitage mass analysis based upon size grades of debitage populations is shown to be prone to errors when making interpretations about the kind of tool produced or the kind of lithic reduction technology used. Significant sources of error may originate from differences in individual flintknapping styles and techniques, raw material size and shape variants, and mixing of debitage from more than one reduction episode. These sources of error render debitage Mass Analysis ineffective for determining the kind of stone tool reduction activities practiced at excavated sites. Mass Analysis may be effective for determining artifact reduction sequences if it is used on debitage from a single reduction episode or part of a reduction episode. However, it is shown that Mass Analysis when used for assessing reduction sequence information, must also control for the effects of raw material variability, assemblage mixing, and flintknapping styles.

Keywords: Lithic analysis; Experimental archaeology; Mass Analysis; Debitage; Size grades

Lithic artifact debitage may be defined as the by-product of stone tool production and resharpening. Debitage includes the chips and debris that have been removed to shape and maintain stone tools. In 1972, Don Crabtree proclaimed that debitage composed the “finger prints” of stone tool production [33]. He rightly meant that debitage could be used to recognize stone tool production activities even in the absence of the stone tools themselves being recovered. Since then debitage has been effectively analyzed to determine the types of artifact produced [15,23,66,67,74,95], the kind of technology practiced [5,30,35,36,54,58,72], the stage of tool reduction or production [8,26,50,61,65,76,85], and even the process of site formation [31,39,49,69,87,88].

The analysis of lithic debitage has become increasingly important in understanding the activities and tasks that have taken place on the prehistoric landscape [10,16,57,59,73,92,93,94]. This has helped researchers understand how human mobility and land-use are linked to foraging strategies, and also how these foraging strategies are linked to lithic tool production, use, maintenance, and discard [14,20,21,32,34,38,47,56,63,91]. A form of debitage analysis known as Mass Analysis (MA) has been widely adopted by archaeologists to process large quantities of artifacts to make such inferences [1,17,18,29,51,62,77,81]. Part of the reason MA has been so appealing to researchers relates to its relative ease, speed, and reliability. The advantages of MA over individual specimen analysis are highlighted by Ahler [1]. He notes that MA:

1. can be applied to the full range of debitage without regarding fracture or completeness, thereby eliminating potential bias resulting from exclusion of some debitage forms, such as broken or shattered pieces;
2. can be rapid and efficient, even for extremely large debitage samples, because it does not require handling and measuring of individual specimens;
3. can reduce technological bias based upon debitage size because different mesh sizes capture a range of different specimen sizes; and
4. can be a highly objective technique since the analysis involves size grading, counting, and weighing, and can be
conducted by virtually anyone trained in elementary lab procedures.

There are indeed appealing reasons for archaeologists to adopt MA. Briefly saying, MA is a form of debitage population study that assimilates all debitage within a recognized population and segregates it into size groups known as “size grades.” Based on the relative proportion of debitage within each size grade generalizations are made about the technology used to produce the debitage population. The relative proportion of size grades is established by the count and weight of specimens in each size grade. The average weight of each size grade and the percentage of specimens with cortex in each size grade may also be calculated. These proportions and average values become the attributes of the debitage population used in MA techniques.

Inferences about the debitage population are generated by a “control group” of debitage based upon experimental replication of various tools or stages of tool production and core reduction. For instance, the investigator might first replicate a projectile point from a cobbles or a flake blank. The debitage from that replication event is sorted into size grades and relative amounts of each size grade are calculated based on counts and weights and/or cortical representation. This control group is summarized to produce a signature of some type such as a histogram, ratio measure, or a discriminant function. This control group signature is then compared to the signature obtained from the excavated collection using the same size grades. If there is a match the investigator may infer that a projectile point was manufactured at that location, even if one was not found there.

A series of replications are then conducted to evaluate all sizes of debitage resulting from the production of different kinds of stone tools and stages of stone tool reduction. These experimentally derived debitage signatures are then used as a reference library for various kinds of tool production or core reduction activities. For example, a specific debitage signature has been produced for hard hammer reduction of cobbles, for bipolar reduction of pebbles, for projectile point manufacture, and even for the initial thinning of bifaces from flake blanks [1,18,68].

The size grades used in MA are easily applied to an assemblage of debitage. One of the most popular ways to stratify debitage into size grades is to sift the debitage through a series of nested screens. The investigator can save a considerable amount of time using this technique because she or he is not required to separate the debitage into whole or broken flakes or shatter or platform remnant flakes. All debitage is shifted through the screens regardless of technological variety or completeness to arrive at size groups or size grades. A fairly untrained technician can do this sorting. In fact, some laboratories stack the nested screens on a mechanical shaker machine so that the sorting or size grading of debitage requires no training at all.

A classic MA study might produce a distribution of debitage size grades like the one shown in Table 1. In this case each of the reduction techniques would have eight variables based upon the proportion of each size grade calculated by weight and then by count. The percentage of cortical specimens for each size grade would add an additional four variables that can also be used in a MA study. These data might then be analyzed to produce a histogram signature that looks like the one in Figs. 1 and 2 showing the relative proportions of count and weight, respectively in each size grade for each reduction technique. In addition, a discriminant function can be calculated for each experimentally replicated assemblage and that signature can be compared to the signature obtained for the excavated assemblage to make an interpretation [1,41].

In this paper, I contend that MA, although it is relatively easy to apply and can be used to process very large quantities of debitage, has been adopted by many practitioners without stringent analytical constraints and with little concern for the validity of the results. Instead, I suggest that this popular analytical technique has been adopted because of its economic benefits (ease and speed) and has unfortunately or unintentionally often resulted in spurious interpretations of the archaeological record. Below, I review several factors responsible for the failure of MA, including problems with replicated assemblages, raw material influences, and mixed archaeological assemblages. I also identify effective applications of MA and

<table>
<thead>
<tr>
<th>Reduction technique</th>
<th>Percent weight</th>
<th>Percent count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Size 1</td>
<td>Size 2</td>
</tr>
<tr>
<td>Hard hammer flakes</td>
<td>29.0</td>
<td>43.5</td>
</tr>
<tr>
<td>Hard hammer blades</td>
<td>17.5</td>
<td>53.0</td>
</tr>
<tr>
<td>Bipolar chips</td>
<td>12.5</td>
<td>34.8</td>
</tr>
</tbody>
</table>

Table 1: Relative proportions of size grades for core reduction techniques by weight and count.

**Fig. 1.** Histogram of relative proportions of count for Mass Analysis of three reduction types (hard hammer flakes, hard hammer blades, and bipolar debitage). Data are taken from Ahler [1].
suggest techniques that can help to make this technique a more productive analytical strategy, such as segregating debitage into discrete technological production episodes.

Before discussing sources of debitage variability that may influence results of MA it is important to clarify how I use the terms "production" and "reduction." I use the term "production" when talking about the manufacturing of "tools" using pressure or percussion flaking methods. I use the term "reduction" when talking about the removal of detached pieces from cores. Reduction refers to the process of flake removal for the acquisition of usable detached pieces. Production refers to the process of flake removal for the purpose of making, reshaping, or reshaping a tool. I refer to core "reduction" and to tool "production." These are not necessarily interchangeable terms.

1. Replicator variability

One of the assumptions of MA is that the production of different kinds of tools or the recognition of tool type stages result in different amounts of debitage in various size classes. It is also assumed that the production of the same kinds of tools by different individuals will result in the same relative amounts of debitage in each size grade. In other words, if two individuals each manufacture a projectile point the debitage produced by each person should result in relatively identical distributions of debitage size grades and ultimately obtain the same signature. A corollary assumption is that excavated debitage matching the size grade distributions of the experimentally replicated debitage will have resulted from the same kinds of artifact production events. Practitioners of MA assume that the debitage signatures produced by contemporary knappers will be the same as the debitage signatures produced by ancient knappers, provided that the same artifacts or the same reduction technologies are replicated.

Is this an accurate assumption? In an independent study of flintknapping experiments, Redman [80] compared the debitage produced from three different "expert" flintknappers making the same type of bifaces using the same raw material and the same kind of percussors. Although she did not evaluate size groups or size grades, she found that there was a significant difference in six debitage characteristics among the experimental assemblages even though all knappers were making the same tool type. As a result of being made by different individual knappers, the experimental debitage was significantly different in flake curvature, flake length, maximum width, width at midpoint, platform width, and flake weight. These data suggest that individual flintknapping abilities, styles, or techniques may produce significantly different debitage size grades. Other researchers have made similar claims [43,70,89]. Individual variation in tool manufacturing has obvious implications for MA, which assumes that replicated debitage assemblages can be used to match excavated assemblages to make inferences about the kind of production that has taken place at a site.

Individual variation in the debitage characteristics noted by Redman [80] could influence the relative proportions of size grades for each debitage assemblage. I tabulated size grade data from several independently published sources and also conducted additional replication experiments to evaluate this possibility. Fig. 3 shows these tabulated data as relative proportions of weight for four size grades produced from bipolar technology [1,52,68]. Morrow’s data were restructured from his original presentation to conform with Ahler’s [1] strategy for naming the largest size-graded debitage “size grade 1” and the smallest size “size grade 4.” Even relatively simple technology such as bipolar reduction produces significantly different debitage signatures when using size grade analysis.

![Fig. 2. Histogram of relative proportions of weight for Mass Analysis of three reduction types (hard hammer flakes, hard hammer blades, and bipolar debitage). Data are taken from Ahler [1].](image-url)

![Fig. 3. Comparison of individual flintknapping debitage by relative proportion of weight for bipolar reduction.](image-url)
For example, Kalin’s experiment produced no size grade 1 debitage while all three other knappers produced at least some of this largest size debitage. In fact, Morrow’s experiment and my own produced most debitage in size grade 1. The highest frequency of debitage from Ahler’s bipolar replications were in size grade 3 while Kalin’s, Morrow’s and my size grade 3 debitage was represented as 5.2%, 13.2%, and 10.2%, respectively. Table 2 lists the significance values for t-tests of all four size grades and demonstrates that the variability in size grade values is too great for discriminating reduction technology using MA, even though each knapper reduced a nodule using the same reduction technology. These data show that each replication could not have been drawn from the same population (have the same manufacturing technique). However, we know this to be the case as all replications were the same — bipolar core reductions. This tends to support Redman’s [80] findings that one of the greatest sources of debitage variability comes as a result of individual knapper’s style and technique.

In a separate study, Amick and Mauldin [4] compared flake breakage frequencies in their experiments against data on breakage compiled by Bradbury and Carr [25]. They found highly significant differences in flake breakage attributable to individual knapper reduction practices for both chert core reduction ($X^2 = 54.35, df = 3, p < 0.000001$) and for chert bifaces/tool reduction ($X^2 = 65.25, df = 3, p < 0.000001$). Again, debitage variability is shown to originate with individual knappers. As such, it is doubtful that contemporary knappers would produce the same debitage size grade signatures as ancient knappers even if both were making exactly the same kinds of tools. Such discrepancies make it extremely difficult to trust results of MA given that MA results are based upon the comparisons and matching of experimentally generated debitage assemblages to ancient debitage assemblages.

2. Raw material size, shape, and composition

Variation in debitage size grades resulting from different flintknappers is only one potential source of error when using MA techniques. Variability in blank size and shape as well as raw material type will also produce different debitage sizes [2,24,96]. However, this fact appears to be ignored by most practitioners of MA. Seldom (if ever) are tool-stone sizes and shapes accounted for when comparing replicated assemblages to excavated assemblages. How can we reasonably expect to obtain the same proportions of debitage weight in each kind of tool-stone, to make the replications and to make the excavated collections, is composed of different sizes and shapes? One way to evaluate this question is to reduce different objective pieces of different sizes and shapes using the same reduction techniques and then compare the debitage size grades using MA.

I conducted such a test by reducing two water worn obsidian pebbles using bipolar reduction technology. Bipolar reduction technology was used because it is a fairly elementary and uniform technique to open tool-stone. Parry and Kelly [73] state, “Cores are not preformed or prepared in any way. Instead, they are struck almost randomly, shattering into pieces of variable size and shape.” Other investigators support this characteristic of bipolar technology [9,44,45,48]. For my replications, Pebble #1 was oval and flat on one side with a maximum linear dimension of 6.2 cm and a weight of 117 g. Pebble #2 was more rounded or egg-shaped with a maximum linear dimension of 8.8 cm and a weight of 339 g. Pebble #2 had approximately twice the mass as Pebble #1, but was only about 30% (2.6 cm) longer in maximum linear dimension. The pebbles were opened in an effort to obtain as many large usable pieces as possible. Table 3 lists the debitage size grade frequencies for each pebble in standard hardware cloth square size increments by percent count and percent weight. The smaller of the two pebbles (Pebble #1), of course, did not produce any pieces in the largest size grade (2-inch mesh size). Pebble #2 only produced four debitage pieces in the 2-inch size grade. However, these four pieces represented 62.5% of the total debitage weight for Pebble #2. This trend immediately shows how different original tool-stone package sizes can influence a size grade distribution used in MA.

The 2-inch mesh size was eliminated from the study to give Pebble #1 and Pebble #2 the same number of size grades with debitage representation (bottom, Table 3). This still resulted in very different relative percentages of debitage weight in each of the four remaining size grades even though both pebbles were opened using the same technology.

Table 3

<table>
<thead>
<tr>
<th>Size grade mesh (inch)</th>
<th>Percent count</th>
<th>Percent weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pebble #1 with five mesh sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td>71.4</td>
<td>4.7</td>
</tr>
<tr>
<td>1/4</td>
<td>18.5</td>
<td>11.9</td>
</tr>
<tr>
<td>1/2</td>
<td>8.3</td>
<td>62.0</td>
</tr>
<tr>
<td>1</td>
<td>1.8</td>
<td>21.0</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Pebble #2 with five mesh sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td>70.1</td>
<td>3.1</td>
</tr>
<tr>
<td>1/4</td>
<td>20.5</td>
<td>8.5</td>
</tr>
<tr>
<td>1/2</td>
<td>5.7</td>
<td>8.7</td>
</tr>
<tr>
<td>1</td>
<td>2.3</td>
<td>17.2</td>
</tr>
<tr>
<td>2</td>
<td>1.3</td>
<td>62.5</td>
</tr>
<tr>
<td>Pebble #2 with four mesh sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1/8</td>
<td>70.1</td>
<td>3.1</td>
</tr>
<tr>
<td>1/4</td>
<td>20.5</td>
<td>8.5</td>
</tr>
<tr>
<td>1/2</td>
<td>5.7</td>
<td>8.7</td>
</tr>
<tr>
<td>1</td>
<td>3.7</td>
<td>79.7</td>
</tr>
</tbody>
</table>
Other investigators studying the influence of lithic raw material on tool manufacturing process have noted similar results. Bradbury and Franklin [27] explored the effectiveness of MA as it relates to raw material package sizes. They conclude that the “...raw material influences appear to be most significant when perceived in terms of the variability in initial nodule size and shape...” [27]. In a test of flake breakage and ultimately flake size, Amick and Mauldin [4] state “...breakage patterns fail to behave consistently with any variables other than raw material...raw material considerations must provide the backbone of any debitage analysis.”

Clearly, raw material composition and package size are important factors influencing debitage size grades regardless of the kind of manufacturing technology practiced. Replication experiments conducted to produce controlled debitage data sets for MA must begin with the same tool-stone configurations as the excavated assemblages to make reliable control groups.

3. Debitage mixing

The most widely recognized problem with using MA is known as the “mixing problem.” Simply stated, if we assume that debitage produced by the production of similar tools results in similar size grade signatures, and we assume this is true regardless of who made the tools, and we further assume that raw material size, shape, and composition are not an issue, MA must still overcome the problem created when debitage from two different production events are mixed together, as is often the case with archaeological assemblages.

For instance, when debitage from the reduction of a biface is mixed with debitage from the reduction of a platformed core, the size grade signatures are no longer discernable or diagnostic. This problem has been recognized over and over again [1,2,30,68,82], but few have attempted to control for it. Even in a controlled reduction context, we cannot use mixed debitage to identify specific reduction activities. In other words, even when we know we are mixing together debitage from blade and bipolar core reduction techniques it is not possible to recognize these two reduction technologies using MA. This problem is certainly magnified when we begin looking at excavated palimpsest assemblages where many different kinds and combinations of tool production and resharpening activities might have taken place over varying periods of time. In an archaeological context, we may not have the relatively simple mixing of two sets of debitage from two production episodes. Archaeological debitage assemblages may represent several complete or partial production events of various tool forms that may have been deposited at different times and in varying amounts.

One way to assess the effects of debitage mixing is to obtain size grade signatures for two different reduction strategies and then compare those signatures to signatures obtained when the debitage from both reduction strategies are combined. I replicated debitage from the production of a large side-notched hafted biface and the reduction of a bipolar core. Both were made from the same kind of obsidian. Fig. 4 shows MA histogram signatures for the production of the hafted biface and bipolar debitage, and demonstrates how those signatures are complicated when those two separate production events are combined. We no longer obtain diagnostic signatures when differential production data are mixed. This is an interesting comparison because the replicated hafted biface began as a flake blank weighing 45.5 g. The final product weighed 32.6 g, and there was only 12.9 g of debitage produced during the production process, which sorted into the two smallest size grades. This created quite a different signature than the bipolar core reduction, which resulted in 168 g of debitage across four size grades. Fig. 5 diagrams this data as linear graphs and shows that bipolar reduction is potentially present, but hafted biface technology is not recognized in the size-graded debitage even though the replicated assemblage contains this reduction. This underscores one of the most important findings of Bradbury and Franklin’s MA study: “…if the experimental assemblage being used to classify other data sets lacks all of the reduction types present in the assemblage being classified, correct classification rates may drop significantly” [27]. We can add to this and also say that correct classifications decline when experimental assemblages contain more reduction types than comprising the excavated assemblage.

4. Diagnostic signatures

There has been an assumption made in the archaeological literature that MA does work in a controlled experimental context. This has been established many times by researchers making replicated debitage assemblages [2,3,17,30,74]. Even when there are problems attempting to match replicated assemblages to excavated assemblages, there appears to be good results for obtaining diagnostic signatures for the
replicated debitage populations. In other words, investigators tend to produce diagnostic signatures for experimental production events using size-graded debitage produced in those manufacturing events. Intuitively, this makes perfect sense because experimentally replicated assemblages are not mixed or subjected to the same type of accumulation problems as excavated assemblages. Nor do replicated assemblages have the problems associated with variation in raw material varieties because this variable can be controlled for in experiments. Unfortunately not all experimentally replicated debitage assemblages produce diagnostic signatures for different production events and technologies.

Toby Morrow [68] produced debitage counts and weights from the replication of a bifacial core, a Clovis point, a blade core, and a bipolar core. Again, I converted his size grades to fit Ahler’s size grade conventions with size grade 1 as the largest and size grade 4 as the smallest. Fig. 6 compares his data as relative frequencies of debitage counts by MA size grades. The debitage from each of these production technologies have almost identical signatures. The highest percentage for each of the four production technologies occurs in size grade 4 and the lowest for all production technologies occurs in size grade 2. Size grade 3 contains approximately 15% of the debitage by count and size grade 1 contains less than 10% of the debitage for all manufacturing technologies. The trends in relative distributions of weight for the same tool replications show the same uniform distribution of size grades, as do the counts (Fig. 7). The reduction of four very different kinds of objective pieces has produced identical debitage size grade proportions using both counts and weights. When these data are graphed linearly (Fig. 8), the similarity is even more pronounced.

Table 4 lists student t-scores on a two-tailed test for each size grade and suggests that all size-graded data could have originated from the same reduction technology. Together, these statistics show that the debitage for each replicated assemblage could have been drawn from a single population. However, we know that was not the case. Four very different kinds of production activities produced each of the replicated assemblages. Mass Analysis of size grades does not effectively discriminate amongst these four assemblages. This certainly suggests serious problems with using MA as a technique for interpreting excavated assemblages of debitage.
5. Reduction stage analysis

Investigations aimed at assessing MA have been very generous in pointing out the usefulness of this method, particularly in conjunction with other kinds of debitage analysis. Most of these investigators suggest that MA is most effective when predicting lithic reduction stages in an experimental context [25,30,68]. This makes some sense because lithic tool production and core technology are reductive processes [12]. Regardless of the kind of core reduction, lithic technology results in detached pieces that get progressively smaller as the objective piece is worked. This trend is clear in the experimentally produced debitage from Morrow’s [68] production of a Clovis point. Fig. 9 shows a scatter plot of the mean flake weight for each of the eight reduction stages in the Clovis point production sequence. Clearly the trend in mean flake weight gets progressively smaller as the production becomes more complete (mean flake weight $y = -0.3626x + 2.8326$, $r^2 = 0.7631$).

An ANOVA shows this to be significant ($F = 19.325$, $p = 0.005$).

However, inferring reduction sequences using MA is also sensitive to the same kinds of issues noted above regarding individual knappers, raw material package size and type, and the “mixing problem.” For instance, reduction sequence debitage from projectile point manufacturing will not necessarily be analytically visible if it is mixed with reduction sequence debitage from a bifacial core reduction or platformed core reduction simply because early stages of projectile point production data may be the same size or even smaller than later stages of core reduction debitage.

Given the information presented above, MA used to determine reduction stages should be cautiously used to infer reduction sequences and only after the debitage assemblage is assessed or stratified to account for factors such as mixing and raw material variability [11]. Making inferences about reduction sequences using MA is only effective if the debitage assemblage represents a single reduction technology or it is segregated to represent a single technology. For instance, it might be effective for determining the kinds of production techniques that have occurred at a small single use or special use camp, such as a butchering station or bivouac, where points were repaired or resharpened, or at a quarry location where cobbles were primarily tested. As the number of production activities and by extension, length of occupation increases at a site location the more problematic MA gives the complications noted above. Also, some types of reduction technology do not result in progressively smaller detached pieces throughout the reduction trajectory. Experiments and analyses conducted on blade and microblade manufacturing have shown that detached pieces from prepared blade cores tend to stay approximately the same size throughout the production cycle [22,37,79]. This is because blade core reduction is selected as a strategy to purposefully obtain regular and uniform detached pieces. These kinds of technologies will not be sensitive to reduction sequence assessment using MA techniques.

<table>
<thead>
<tr>
<th>Size grades</th>
<th>One-sample test</th>
<th>Mean difference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$t$</td>
<td>df</td>
</tr>
<tr>
<td>Size grade 1</td>
<td>21.76</td>
<td>3</td>
</tr>
<tr>
<td>Size grade 2</td>
<td>5.13</td>
<td>3</td>
</tr>
<tr>
<td>Size grade 3</td>
<td>4.71</td>
<td>3</td>
</tr>
<tr>
<td>Size grade 4</td>
<td>3.49</td>
<td>3</td>
</tr>
</tbody>
</table>
6. Discussion

Various forms of MA have been used by lithic researchers in the past [3,6,7,19,46,74,75,78,93], however, it did not become popular or widespread as a methodology until Ahler published his 1989 version. Ahler was very careful in applying MA to his assemblages of debitage. He anticipated and discussed many of the potential problems with using MA cited here, such as the mixing problem and issues with raw material variability. Unfortunately many practitioners of MA were not as careful as Ahler when adopting it for their own research. I do not believe Ahler would approve of some of the applications of MA evident in the literature today.

Mass Analysis may be effective for some types of assemblages and for gaining insights into some questions and technological issues. However, the reasons for why it was originally adopted and has become widely used may no longer be applicable to this technique. Given the need to stratify debitage assemblages to counter some of the effects of raw material variability and mixing technologies before conducting MA, it may not be as time conservative as once thought.

This paper did not explore and evaluate the different techniques available to obtain debitage size grade signatures such as discriminant function analysis [1], multiple linear regression analysis [82,90], or fractal analysis [28]. Instead, this paper focused on the issues surrounding the production of primary comparative assemblages (replicated assemblages). Without reliable comparative assemblages, interpretations of size grade debitage are useless regardless of the technique administered to derive the signatures used to compare the excavated assemblage and the replicated assemblage.

It is becoming increasingly clear that debitage assemblages should be stratified into as many behaviorally meaningful groups as possible before MA is performed [11,83]. For instance, it makes little sense to use MA on aggregate assemblages comprised of different types of raw materials or variations of the same raw material type. These different raw material packages most likely represent different reduction or production episodes, even if the materials were used to make the same kinds of tools [53,59,60,86]. The debitage from each of those episodes should be handled as separate analytical populations for MA to be an effective form of aggregate analysis.

In addition to stratification by raw material varieties, evidence for different kinds of lithic production technologies can be used to stratify debitage assemblages to help reduce the amount of assemblage mixing before MA is preformed. There is no reason to expect that MA will produce accurate results if the debitage from multiple production technologies are combined. Debitage with characteristics diagnostic of specific production/reduction technologies (such as bifacial trimming flakes, notching flakes, prismatic blades, bipolar flakes, etc.) should be partitioned into separate analytical populations before attempting MA. For instance, if we are able to identify debitage associated with bifacial production or bipolar reduction, or platformed core reduction, this information can be used to help stratify debitage assemblages into specific production episodes to reduce the mixing problem so often associated with MA. Such technological information is available to debitage analysts [12,33,40,42,44,55,64,71,83,95,97].

Because size grade corresponds to uniform and universal mesh dimensions, many believe MA is a more objective means of classifying debitage by size [1]. We now realize that size grades that correspond to mesh sizes may not be as reliable or as objective as once believed [13,83,84]. When debitage is sorted and passed through the mesh it may get trapped in the mesh in an unsystematic manner. The same debitage assemblage might have different group representation if it is sorted multiple times simply because of how individual flakes are oriented when passing through the mesh. Long thin flakes may pass through a particular mesh size on one occasion but not the next time simply given the flake orientation. Carr and Bradbury [30] suggest passing specimens through a screen by hand to circumvent this potential problem, or passing debitage individually through size templates [12].

A recent example underscores how MA can lead to spurious conclusions. Franklin and Simik [41] conducted artifact-refitting studies and compared their results to the results obtained using MA based upon discriminant function signatures on principal components of size grade attributes. Their refitting study conclusively established bipolar techniques as the primary technology used to open and to reduce chert nodules at a rock-shelter in Tennessee. A series of cobble reduction experiments were conducted to establish comparative debitage collections (of the same raw materials) for bipolar reduction, hard hammer cobble testing, hard hammer core reduction, and soft hammer tool production. When these control populations were compared to the excavated assemblage using MA, bipolar reduction was not recognized. Again, even though bipolar technology was conclusively recognized by refitting cores, MA was not effective in discerning this manufacturing technology. This is not surprising given the complications of performing reliable MA as noted above.

There is another issue with MA related to the circularity in its logic. The potential benefit of using MA for investigators lies in its ability to make inferences about the kinds of tools produced or the kinds of reduction activities that have taken place where debitage is found. However, for MA to be effective a replicated assemblage must be produced to compare with the excavated assemblage. How does the investigator determine the kinds of tools or technologies to replicate? This is a serious circularity problem. If the investigator knows the kinds of tools made to produce the excavated debitage assemblage — why conduct a MA if the aim of MA is to determine the kinds of tools made? If the investigator does not know the kinds of tools made to produce the excavated debitage — how can reliable replications of the assemblage be conducted, particularly given the fact that tool-stone size and shape, as well as production activities, influence the size and amount of debitage produced.

7. Conclusion

The results of this study emphasize that size grade patterning of debitage used for MA is subject to multiple
slopes of error. Production debitage is greatly influenced by replicator variability and it is unlikely that contemporary experimental assemblages can be expected to match excavated assemblages, even if the same types of artifacts were manufactured. Individual knapping styles have as much influence on debitage size as does the kind of tool produced. I also demonstrated how raw material variability, particularly package size and shape, could also influence the size of debitage produced in replication experiments. Raw material type and form must be accounted for when comparing excavated assemblages to replicated assemblages by size grades. Experimentally replicated debitage that has been mixed does not accurately reveal the technological activities from which the debitage was produced. This problem is almost certainly more pronounced when dealing with excavated assemblages where multiple lithic production activities may have been performed. Lastly, I showed how some replicated production activities do not produce significantly different debitage assemblages when making size grade comparisons, even when very different kinds of tools are produced. All of these potential and probable sources of error make using MA a questionable analytical strategy for interpreting excavated debitage assemblages unless these factors are adequately controlled and accounted for.

Clearly, MA of experimentally generated debitage assemblages has been an effective heuristic for chipped stone technological analysis, even if they are difficult to confidently relate to excavated assemblages. We have learned a great deal about the sources of debitage assemblage variability as a result of our explorations into MA. However, as shown above, MA is not effective for making accurate tool production or core reduction interpretations at archaeological sites unless we can account for the many sources of debitage size variability.

Acknowledgments

A version of this paper was presented at the 2005 meeting of the Middle Atlantic Archaeological Conference in Delaware. I thank those organizers and colleagues for providing a forum and comments to that initial draft. The College of Liberal Arts at Washington State University provided professional leave time and funding through their Edward R. Meyer Project Grant to allow me to complete this paper. I am grateful to them and appreciate their contributions. A number of colleagues have provided valuable editorial comments and suggestions to the early versions of this manuscript. I acknowledge the helpful commentary of four anonymous JAS reviewers and Bill Lipe, Jennifer Keeling, Beth Horton, and especially Karen Lupo. All have contributed towards helping my less-than-concise manner of writing.

References

[2] Stanley A. Ahler, R.C. Christensen, A Pilot Study of Knife River Flint Procurement and Reduction as Site 32DU508, A Quarry and Workshop Location in Dunn County, North Dakota, Department of Anthropology and Archaeology, University of North Dakota, Bismarck, 1983.


[59] Mary Lou Larson, Site formation processes in the cody complex at the laddie creek site (48BH345), Wyoming, Geoarchaeology 7 (1992) 103–120.


[61] Mary Lou Larson, Judson B. Finley, Seeing the forest but missing the trees: production sequences and multiple linear regression, in: Christopher Hall, Mary Lou Larson (Eds.), Aggregate Analysis in Chipped Stone, University of Utah Press, Salt Lake City, 2004, pp. 95–111.


