Designing and Implementing a System for Assigning Student Identifiers

This paper identifies and discusses issues related to the creation and maintenance of a student identifier system. Options for designing that system and recommendations for consideration are presented.

Prepared for the

Massachusetts Department of Education

by

Glynn D. Ligon, Ph.D.

Evaluation Software Publishing, Incorporated

December, 1997

Two related papers will follow on the results of a national survey of state education agencies and their policies and practices related to student identifiers, and issues related to the confidentiality of student information in statewide databases.

Evaluation Software Publishing, Incorporated
1510 West 34th Street Suite 200
Austin, Texas 78703
512-458-8364 Fax 512-371-0520
esp@evalsoft.com www.evalsoft.com

Designing and Implementing a System for Assigning Student Identifiers

Glynn D. Ligon, Ph.D.
Evaluation Software Publishing, Incorporated

This paper identifies and discusses issues related to the creation and maintenance of a student identifier system. Options for designing that system and recommendations for consideration are presented.

December, 1997

Imagine high school lockers without a number on the front of each one. Students could find theirs by remembering location relative to hallways and windows, but the assistant principal would have to open them up and examine contents to know whose was whose. Student databases are very similar. We can put data into records (lockers) for each student, but we need an identifier (or number) to locate the record quickly for a specific student. The alternative is to examine the contents of each record each time to know whose record is whose. Imagine if an individual student's possessions were distributed across numerous lockers. With locker numbers, we can find all those possessions efficiently. If all of the student's lockers had the same number, we could find them even more efficiently.

With identifiers, we solve the problem of how to index pieces of information about the same individual spread across separate files and tables within the same file, so they can be linked into a complete set of information about that individual. The new problem they create is the added burden of assigning, maintaining, and entering those identifiers into all records. Large-scale databases could not function efficiently if they had to use names and other pieces of information to link pieces of an individual's records to each other. The burden imposed on us to remember and record our identifiers is easily justified by the benefits and efficiencies gained within our information systems.

State education agencies (SEAs) find the need for student identifiers growing with the increasing demands for a central information source for decision making. In response to this, SEAs are developing more automated information systems. These software applications want information to be codified to save space and to allow for linking of information across separate files. Although some information systems use only the student's name, those that rely upon related information in other files, longitudinal comparisons, or efficient searching for records have an identifier system.

Today, the burden of managing identifiers has been heightened by the necessity of accommodating legal and emotional demands for confidentiality. The identifier itself has become the object of confidentiality concerns, because the identifier becomes a virtual key to unlock access to personal information. In some cases, the identifier itself contains

imbedded information about the individual. A discussion of student identifiers cannot be conducted in isolation from associated confidentiality and security issues.

However, within this challenging context, local education agencies (LEAs) and SEAs have developed well-designed and managed student identifier systems. None was developed without influence by political and legal forces.

There is a very practical side to student identifiers. These identifiers become the index used by a computer program to find an individual's record(s). In a relational database, the identifier takes on a heightened sense of importance because information about an individual is stored in separate tables rather than in a single, long string of fields.

Some may expect that with an adequate student identifier there would no longer be a need to store a student's name in a database. The reality is that even the best identifier needs to be verified. Verification means having other pieces of information available to determine that the identifier is attached to the correct student's records. This verification does not have to occur within the main database, but somewhere there must be a table that matches the identifier with other data elements such as name, birthdate, and gender to allow periodic verification.

This paper will describe and discuss issues related to the creation and maintenance of a student identifier system. Options for designing that system and recommendations for consideration are presented.

Defining Terms

The definition of terms as they are used in this paper is important to facilitate a precise communication of the issues and recommendations described.

Identifier: A set of characters used as a proxy for name to identify uniquely an individual (Identification number is not used because that term denotes a numeric identifier. ID is not used because it at times refers to the physical medium, e.g., card, used to document an identifier.)

Registrar: Any school representative who registers a student; the person who collects registration information from a parent or student, including the securing of an identifier or the assignment of a new one (In this use, registrar is a role or activity, not a job title.)

Registration: The provision by an individual of personal information to establish eligibility for enrollment (Registration may occur at a central site or at the school building. A student may be registered but not yet enrolled in a school.)

Enrollment: The listing of an individual on the membership roll of a school for the purposes of attendance

Access: Viewing or gaining possession of information or data elements from a database or other product of an information system

Secure: Describes information and data elements protected from unauthorized access by processes that manage access

Confidential: Describes information and data elements that must not be accessed by persons without the authority to know them

Directory Information: Data elements not designated as confidential by law, regulation, policy, or practice (e.g., name, birthdate, school, grade level, gender, place of birth, parents' names, address, phone number, etc.)

Duplicated: When an identifier is assigned to more than one individual

Alias: A prior identifier that is not used when a new identifier is assigned to the same individual

Unique: When an identifier appears only once in the pool of valid identifiers and is assigned to only one individual

Permanent: When an identifier is used for an individual as long as any record of that individual is maintained in the database

Ubiquitous: When the same identifier or identifier system is used whenever and wherever data about an individual exists throughout an entity's information system

Nominal: When an identifier contains no intrinsic or embedded meaning other than to designate an individual (An encrypted identifier is not nominal because meaning is imbedded in its characters even though that meaning may be disguised.)

Assignment: Linking an identifier to an individual

Verification: Confirming that an identifier is correct for an individual

Correct: An identifier that is the actual and legitimate identifier for an individual

Valid: An identifier that is from the authorized pool of identifiers and meets all criteria for being an identifier

What is the Nature of a Functional Student Identifier?

What is the essence of a student identifier? A functional student identifier must:

1. Be a stand-in for a student's name. Why?
    a. Names change, their components change order, they are abbreviated, they are not reliably provided by a student or parent from one time to the next.
    b. Names are of varying length, so a file has to allocate long fields that take up substantial space in a database.
    c. At times we do not want a student's name showing on a document.

2. Be unique, unduplicated, and permanent. Why?
    a. Students move, change names, go away and return, but databases must be able to link a student's records across time, across files, and across schools.
    b. The identifier must be unique to identify one and only one student or else records could be inappropriately combined for more than one student.
    c. The identifier must not be an alias because the student must have one and only one number, so all of the student's records can be linked into one set.
    d. The identifier must be permanent because changes in a student's identifier impose extreme challenges to link multiple identifiers to the same student.

3. Be ubiquitous. Why?
    a. The records and systems that use the same identifier can be linked to create a set of records for an individual student. (Note that crosswalk tables can be used to translate the identifier used by one system to the identifier used by another. However, creating and maintaining crosswalk tables is difficult and adds a step to all uses of the data.)
    b. Records that do not use an identifier must be linked using less precise matching techniques.

The bottom line for a student identifier is that it must identify one and only one student "forever." Forever is defined as being as long as records for a student are maintained within the database—even if the student leaves the state for a time and returns later.

Questions the SEA Must Answer to Define the Appropriate Student Identifier

There are some basic questions that determine the nature of an organization's identifiers and the system used to assign, verify, and use them.

1. To what level must the identifier be unique?
    a. Schoolwide
    b. Districtwide
    c. Statewide
    d. Nationwide

2. What level of burden can be imposed?
    a. None
    b. Minimal
    c. Moderate
    d. No Limit

3. Where will the identifier be assigned?
    a. At the school level
    b. At the district level
    c. At the state level
    d. At the national level, e.g., Social Security Number (SSN)

4. When will the identifier be assigned?
    a. At the time of registration
    b. Within a few hours of registration
    c. Within a few days of registration
    d. Upon first reporting of the student to a central system

5. At what level will the identifier be verified?
    a. School
    b. District
    c. State

6. How will the school/district investigate the identifier quickly when a student enrolls to verify an existing identifier or to assign a new identifier?
    a. Query parent or student
    b. Contact prior school
    c. Examine transferred records
    d. Connect to a statewide database

7. What information will be used to verify the correctness of an identifier?
    a. Directory information
    b. Detail personal information

8. How public or confidential must the identifier be?
    a. Not used on public documents
    b. Not used locally; used only for state reporting
    c. Public as part of defined directory information

9. How much information can the identifier carry?
    a. None, the number must be merely a nominal identifier.
    b. Some directory information such as district, school, name, birthdate, etc. can be imbedded in the identifier.
    c. Some directory information can be imbedded, but there must be an encryption algorithm to mask the information.

10. How long can the identifier be?
    a. No more than fields in existing databases and software applications
    b. Any length

11. Can the identifier contain nonnumeric characters?
    a. No, too many errors can occur.
    b. Yes, otherwise the identifier will be too long.

12. Will rules be applied to avoid common problems by users of the identifier?
    a. No leading zeros
    b. No more than two consecutive identical numbers
    c. No final zero
    d. Check digit to flag invalid numbers

Options for Selecting a Student Identifier

SEAs have adopted a variety of identifiers. Their approaches and some avoided by them are categorized below along with an analysis of the advantages and disadvantages of each.

1. **SSN:** Use of the SSN as an identifier is legal and in practice in a number of states. With the requirement by the IRS for dependents to have a SSN, almost all students enter school with a SSN assigned. The use of SSN for student identifiers has been a common practice by postsecondary institutions for decades.

    Option 1.a. Require SSN

    Option 1.b. Encourage use of SSN, but provide for an alternate identifier at the parent's request

    Option 1.c. Collect the SSN as a data element, but use another identifier

Advantages: SSN is unique on a nationwide basis. SSN is almost universal in its assignment in the U.S. The assignment and maintenance of the SSN system is handled by the Federal government; thus, the burden of determining an identifier and assigning it is reduced. With SSN, the parent and student can assist in providing the identifier when transferring from one school to another. When students move across states, those states using SSN can use it to verify records. Postsecondary institutions are assisted in their applications processes when secondary schools can provide SSN on transcripts.

SSN can be used to share information or conduct studies across agencies that use it. Some states exchange information about families across agencies to determine eligibility for services. For example, several states use SSN and other family information to link across Aid to Families with Dependent Children and other public assistance files to establish a student's eligibility for the school meals program, to count the number of economically disadvantaged students to qualify a campus for Title I funding, and to establish a student's eligibility for vocational and job training programs.

More controversial is the ability to exchange student information with law enforcement agencies such as the local police or the Immigration and Naturalization Service. Confirmation of a student's identity and enrollment status can assist their investigations. At times, LEAs and SEAs are legally required to cooperate and provide information. Within that context, ensuring that the information provided actually belongs to the correct individual is critical.

SSN is the best identifier for use when conducting former student studies. Employers, the armed forces, and postsecondary institutions can use it to correctly match former students with their current employees or enrollees.

Disadvantages: Some parents are uncomfortable providing SSN, others strongly oppose its use. A very small number of students will not have one. Parents have occasionally provided their own SSN for their children. To use SSN, the SEA would be obligated to provide an alternative identifier to parents or students who refuse to provide it. The requirement to design and implement an alternative identifier is equivalent to having two identifier systems in place. Students may move back and forth between the use of their SSN and their request for an alternative identifier. Tracking these changes in a longitudinal database is difficult.

In states using SSN, an attorney general's opinion, legislative authority, or state board of education authority is typically secured first.

Consideration of SSN adds time and effort to the planning, review, and public comment process.

In contrast to a nominal identifier, the printing or display of SSN on education documents demands a higher degree of diligence from everyone handling those documents.

2. **Local-Assigned Identifier:** Some LEAs have extant identifiers, and some are assigned at the school level by student information management software applications. The SEA could use these pre-existing identifiers.

   Option 2.a. Use pre-existing local identifiers in year one only, then assign all new numbers from the SEA system.

   Option 2.b. Authorize LEAs to continue to assign identifiers as is current practice.

   Option 2.c. Add a district identifier to the front of locally assigned identifiers to make them unique across the state.

   Advantages: Local control is maintained. LEAs are not required to make changes in their identifier system. Historical local files continue to be compatible with the identifiers. An identifier can be assigned at the time of registration for new students.

   Disadvantages: Uniqueness across the state is not likely. Mobile students would receive a new identifier in each district/school. The formats and characteristics of local identifiers would have to be considered in the establishing of parameters for acceptable identifiers. SEA's central database design and the checking for aliases and duplications would be more complicated. 2.c. might work with a district identifier added to the front of local identifiers to make them unique statewide. However, where the local identifiers are only unique within a school, both a district and a school identifier would have to be added. This has implications for the length of the identifier.

   This strategy breaks down when students move from one district to another. Each district would have to accommodate the characteristics of other districts' identifiers or there would be aliases created in the assignment of numbers to individual students. The SEA would have to abandon uniqueness statewide to fully accommodate option 2.b.

   Option 2.a. would require the setting of criteria for a local identifier to be acceptable, e.g., no longer than the SEA identifier, same characteristics in

regard to alpha, numeric, and special characters, etc. This might eliminate too many local identifiers from use.

Option 2.c. works only if new schools verify a prior identifier from a prior district rather than assigning a new local identifier. The addition of a three-character district identifier to the front may make the identifier longer than practical.

3. **Algorithm to Assign Identifier:** An algorithm could be devised that encrypts the student's name, birthdate, gender, place of birth, and possibly other data elements to create a unique combination. The algorithm could be secured to protect the contents of the identifier.

   Advantages: The algorithm could be distributed as a software application to registrars for use at the time of registration. Parents and students would not have to know the identifier nor would registrars have to contact prior schools.

   Disadvantages: The required length of the identifier to ensure uniqueness might be excessive. A hacker (an unscrupulous computer user) might break the encryption routine. Changes or mistakes in the data elements used would result in incorrect identifiers.

4. **State-Assigned Identifiers:** The SEA creates a pool of valid and available identifiers and each student new to the state is assigned a permanent identifier.

   Option 4.a. Identifiers are assigned from a common statewide pool.

   Option 4.b. LEAs are assigned a block of identifiers from the state pool.

   Advantages: The SEA controls the characteristics of the identifiers. The validity of identifiers can be verified by the SEA. Uniqueness is assured. Option 4.b. facilitates assignment of identifiers to new students at the time of registration. The identifier is not directly linkable to confidential data sources.

   Disadvantages: LEAs must rely upon the SEA for their identifiers. Access to their identifiers and the ability to assign them at registration will require sound management.

Analysis of Options

These options cut across issues. In the analysis that follows, the recommendation offered for one option will interact with the recommendation for another.

1. Uniqueness--At what level must the identifiers be unique?

An identifier must not be duplicated within a population or be an alias for a single individual within the population. Uniqueness must be maintained at the state level for a statewide identifier to ensure accuracy and completeness of data. The current identifier assigned by schools and districts to their students may not be unique across other districts. (In fact, some commercial student information management systems only provide uniqueness within a school building.)

Statewide uniqueness can be achieved if districts assign identifiers that are unique within the district and begin with a district number. However, this complicates the process of verifying and using those numbers when students move across districts.

There are certain benefits to using an identifier that is unique nationwide. At this time, the only such identifier is the SSN. SSN provides functionality for tracking former students into postsecondary education, verifying the identity of students across states (with agencies using SSN), and exchanging useful data with other state and Federal agencies that provide services to families and individuals (e.g., verifying eligibility for services). These benefits can be achieved by collecting SSN irrespective of its use as the student identifier.

RECOMMENDATION: Create student identifiers that are unique at the state level.

2. Burden--What level of burden should be imposed upon LEAs?

Burden is defined as the time, effort, and resources required to implement the student identifier system. This includes the necessity of creating the system, assigning the identifiers, verifying an individual's identifier, and entering the identifiers wherever they are required. Burden also includes the effort to make the transition from an existing identifier system to a new one.

Clearly the level of burden must be limited to achieve compliance (both voluntary and practical) with the identifier process. Too high a level of burden will introduce unwanted errors as a consequence of the attention to detail required. Burden must be balanced by benefit. In the case of identifiers, the benefit must be assumed to be high because they are critical to the functionality of the entire student information system. The options that impose the least burden, school and district identifiers, fail to provide the functionality required as described by other

issues. Burden is typically an issue to recognize and to manage rather than being one of the deciding issues when a student identification system is selected.

> **RECOMMENDATION: Accept a moderate level of burden in exchange for the clear benefits from a functional student identifier within the information system.**

3. Assignment--At what level will the identifiers be assigned?

This issue is related to burden. Assignment of the identifiers at the school level provides the quickest and least burdensome alternative. However, this issue is also related to uniqueness. The schools must follow a procedure that ensures unduplicated identifiers. Therefore, identifiers should be assigned at the lowest level possible without losing their uniqueness.

A reality is that parents and students cannot be relied upon to carry their student identifiers from one school to the next. Mobile families too often cannot identify their last school/district, do not have records with them, and cannot remember student identifiers.

> **RECOMMENDATION: Assign identifiers at the level where registration occurs. However, this does not preclude the use of state-assigned identifiers or a student locator system that provides the identifier to be assigned.**

4. Timing--When will the identifier be assigned?

Schools need an identifier immediately upon enrollment of a new student. Certain forms are completed at that time and begin to go their separate ways. Ensuring that the student's identifier is on each form immediately saves changes and mismatches later. The difference between immediately and within a few hours is arguable. However, any system that takes days to assign an identifier presents a very different level of burden.

> **RECOMMENDATION: The identifier should be assigned at registration and be available to schools before enrollment forms begin to be distributed to their respective offices.**

NOTE: SEAs typically require that their official student identifier appear on all reports and data submissions from the LEA. There is usually not a requirement that the state's identifier be used on all local files and records. A district or school could opt to use its own identifier system for local applications such as scheduling and grade reporting. A crosswalk table could be used to translate local identifiers to the state identifiers whenever reporting to the state is required.

5. Verification Level--Where will the identifier be verified?

Verification is the process available to ensure that the identifier assigned to a student is valid and correct. Valid means that the identifier is one actually included in the pool of identifiers to be assigned. Correct means that the identifier is accurately matched to the student.

The identifier can be verified immediately upon registration or later as part of a validation process at the state level. The earlier the verification occurs, the fewer changes will be required later if an identifier is changed/corrected. The closer the verification occurs to the parent and student, the higher the probability of accuracy. Verification at the time of registration, when the parent and student are most likely to be present, is best. This requires that the person registering the student be authorized to assign the identifier according to a set of precise rules, or that the person have direct access to the assignment process.

Verification conducted at the state level after submission of individual records using available demographics in the database is the least efficient. Correct identifiers can be challenged based upon duplications in the data elements used for verification, e.g., students with the same name, birthdate, and gender. In these instances, verification is then delegated back to the school. This state-level verification is a required component of the system, but the frequency of potentially incorrect identifiers can be greatly reduced by adequate controls at registration.

RECOMMENDATION: The identifier should be verified at the time of registration when parents and students are available to answer questions.

6. Assignment/Verification Process—How will the identifier be assigned or verified?

The assignment process includes several steps:

First, the registrar determines if the student already has an identifier assigned.

Second, the registrar secures the existing identifier or causes a new one to be assigned.

Third, the registrar records the identifier in local records.

The registrar can accomplish the first step by asking the parent or student, or examining paper or electronic records from a prior school. Parents and students

too often do not have the records, and at times cannot precisely enough identify contact numbers or addresses for the prior school (e.g., districts with county or descriptive names rather than city names). In the case of migrant worker families, enrollment in a prior school may have been too brief to generate an official record.

An alternative is to establish a statewide reference, such as a web page, that can be queried to determine the existence of a previously assigned identifier. The registrar could access the reference during the registration process. The second step could be accomplished using the reference to learn the identifier or to request assignment of a new one. In the absence of such a reference, the registrar must contact the prior school. This is a critical point. Students who have existing identifiers can be assigned an alias identifier simply because that is easier than contacting a prior school, or because the prior school cannot be contacted or does not respond promptly.

RECOMMENDATION: Registrars should assign and verify identifiers at the time of registration using a statewide reference for finding existing identifiers and a state-approved process for assigning new numbers.

7. Verification Data Elements—What data elements are required for verification of a student's identifier?

When a student's identifier is in doubt, other unique combinations of information about the student must be used for verification. This is typically "directory information" as defined by FERPA along with a few other pieces of information included to increase the probability of describing a single student. The directory information could include name, birthdate, place of birth, gender, prior schools/districts of enrollment, enrollment dates, and grade level. Additional detail information could include parents' names, ethnicity, and SSN.

Security and confidentiality issues must be considered. However, the more information available for query, the more likely existing identifiers will be found and used.

RECOMMENDATION: As many of these data elements as practical should be included in the verification resource.

8. Confidentiality--Who may know the identifier?

FERPA, Massachusetts 603 CMR 23.00, and local policies will define the answer. If the identifier used is SSN, then SS laws would restrict access as well. If the identifier is a nominal code without intrinsic meaning, then it may be viewed as directory information.

Prudent practice would call for the number to be treated as confidential, because knowledge of the number would place the holder one step closer to access to confidential information.

> **RECOMMENDATION: Only education employees with a need to know should access the student identifier.**

9. Imbedded Information--What meaning will be built into each number?

SSN has no important intrinsic meaning imbedded in the numbers. (Generally, the number may imply a region, year of assignment, or sequence, but any algorithm producing the numbers is obscure.)

Imbedded information typically adds to the length of a number. For example, county/district codes add six characters to a number. Birthdate adds six or eight.

A truly unduplicated, random number carrying no meaning has the advantage of requiring less restrictive security and confidentiality precautions.

If the district number is imbedded, this could identify the student's first district of enrollment; however, that information can be carried in other fields within the database as well.

> **RECOMMENDATION: There should be no imbedded information in the student identifier.**

10. Length--How many characters can be in each number?

Shorter numbers can be entered, transcribed, and maintained with fewer errors. A common length provided for an identification number on generic scanner documents is 10. SSN is currently 9, but moving to 10 numbers has been discussed. To accommodate a million students without reusing numbers, requires a minimum of 7 numbers. This provides 10 million minus one unique numbers.

To estimate a maximum quantity of identifiers required, assume that 1 million students enroll in year one, that the identifiers are numeric, and that the identifiers need to be maintained as unique for at least 15 years. Assume a student mobility rate of 20% annually (20% of the students enroll in a new school in a 12-month period), and all of the mobile students leave the state or come from outside the state. Then about 4 million numbers are required.

Student mobility research estimates that half of student mobility is within the same district; half of the remaining half (or 25% of the 20% overall = 5%) is across state lines. Massachusetts may be higher with so many boarder states.

New kindergarten or prekindergarten students could require about 10% new numbers annually. So with 5% for mobile students, 10% for kindergarten students, and a generous 5% for administrative margin, an annual use of 20% new numbers would require only the 4 million numbers across 15 years as described above.

To be prudent, adding additional digits provides a cushion for unanticipated events, i.e., the desire to imbed meaningful information.

An 8- to 10-digit number appears to be a very safe minimum.

RECOMMENDATION: Create a 10-digit student identifier.

11. Characters--What should be the nature of the characters?

Any number, letter, or symbol could be used. Symbols and letters present problems with recognition and accuracy in entering—especially when mixed with numbers. Certain letters (e.g., o, l, i, z, E, b/d, q/p) are sometimes confused with numbers or each other.

Using both numbers and letters provides for many more combinations for unique identifiers, and thus the ability to have shorter identifiers. Problematic numbers and letters could even be eliminated from use (e.g., neither 0 nor o ever assigned). Some state systems use letters and numbers in combination (e.g., Texas starts a state assigned number with S to distinguish it from a SSN.)

Letters require too many bubbles on a scanner form. Letters come in capital and lower case forms that may or may not have meaning, but often cause confusion as to their use. Numbers are easier to distinguish from each other, they can carry intrinsic and extrinsic meaning, and they are more "universal" across languages and cultures. Numbers can be assigned without risking the creation of meaningful and undesirable combinations as with letters.

In the absence of considerations that require restricting the length of the identifier, use of only numerals is preferable. They are easy to distinguish. They can be entered with efficient keystrokes using a number pad. They require less space and are associated with less bubbling error on scanner forms.

RECOMMENDATION: Only numerals should be used for the student identifier.

12. Rubric--What conditions will be imposed on the numbers?

If an algorithm or imbedded meaning is to be used, then those considerations will answer this question. However, if a random number is used, then several rules can be followed to reduce data entry and clerical errors.

Leading and final zeros are sometimes accidentally dropped when numbers are entered. If the remaining numbers are justified left or right, then a reader or a computer application can interpret an incorrect number.

Consecutive identical numerals at times are entered too few or too many times.

Eliminating all the cases described above would reduce the available pool of numbers by about 15%.

A final check digit (a number calculated by formula from the other digits) is sometimes used to provide a quick way to locate invalid numbers. If the formula does not generate the final digit as in the number reported, then there is an error. This check digit could be the 10^{th} character. If a check digit is used, and it is not allowed to be zero, then some of the numbers eliminated because they end in zero would be recovered.

> **RECOMMENDATION: Use unduplicated random numbers with the exception of any with an initial zero, or any sequence of three or more identical numerals. Calculate a final nonzero check digit.**

Attachment A is a summary of the issues and recommendations.

Policy Implications

Several of the issues discussed raise policy questions. In Attachment A, the options that require policy consideration are highlighted. In general, any use of SSN is designated as having policy implications. This designation is not a legal determination, but a practical recognition of the controversy that is associated with the use of SSN. As stated earlier, SEAs using SSN have typically acted under or sought an attorney general's opinion, legislative mandate, or an official regulation.

The level of confidentiality ascribed to the identifier is a direct outgrowth of applicable FERPA and Massachusetts 603 CMR 23.00 provisions. If the identifier contains meaningful elements, even if they are encrypted, a policy decision would be advised to determine if the identifier as a consequence was confidential. Even in the absence of any meaningful content, e.g., a truly random identifier, the identifier might be designated as confidential except for educators and other governmental employees with a need to

know. This designation in part could be based upon the potential for accessing other information using the identifier if it were to be known.

The level of confidentiality would impact the openness of any verification reference. For example, a web page accessible by districts to look up identifiers or to assign an identifier for a new student would need to be compliant with adopted policy.

Associated with the definition of the identifier and its use within the overall student information system, is the issue of defining the basis for "a need to know." This phrase communicates a very general meaning that a person must have a legitimate educational reason to access confidential information. A clear statement of who has a need to know under what circumstances and which data elements they can know is a basic component of a functional student information system. The role of the student identifier should be defined within this statement as well.

In general, an identifier other than SSN would not need a high level of confidentiality as long as prudent firewalls and access procedures are in place to manage access to the databases within which confidential information exists.

Summary of the Recommendations

These recommendations support Option 4.b., the use of a unique, numeric identifier determined by MDE. However, the benefits to adopting SSN as the identifier have been compelling in other states. SSN would be an alternative recommendation if politically acceptable. If SSN is adopted, then the recommended state identifier described here would still need to be used as the alternative number when parents or students decline to provide the SSN. The major factor influencing the recommendation is that Massachusetts will have functional and universal Internet access to MDE from districts. This makes the use of a network-based (e.g., Internet) locator reference practical. Without the locator, verifying identifiers and ensuring the assignment of unique identifiers to new students is more challenging.

If SSN is not used as the identifier, then it should be considered for collection and inclusion in the locator and central database. The added value from having the SSN in a student's record is great.

What design matches the recommendations?

Attachment B is a diagram showing the process of assigning, verifying, and maintaining the recommended identifier system. This diagram also illustrates the relationships among the student identifier, a web-based locator, and the central student database.

The Identifier: The identifier recommended is 10-digit, numeric, and generated as a pool of unduplicated numbers without initial or final zeros, and no digits repeated three or more times sequentially. The use of a check-digit is recommended.

In order to meet the needs of schools wanting an identifier at the time of registration, and knowing that there will be glitches that will at times delay access to a locator system, assigning a block of identifiers from the statewide pool to each district is recommended.

The Central Student Database: This central student database is shown as only very generally representative of the actual database to be created. The central database would not be required to contain the student's name and identifier. In Attachment B, a confidential crosswalk table is shown. This table would translate the identifier reported by the districts to an internal key. The internal key would become a field in the central database as a proxy for name and identifier.

An important point here is that the internal key would be crosswalked, not encrypted. An unauthorized user would not be able to use the internal key to decrypt the student's official identifier.

The Confidential Crosswalk Table: This table receives the encrypted submissions, strips off the identifier and the student name, and replaces them with an internal key (a unique, random number). The decrypted and modified file is then uploaded to the central student database. The confidential crosswalk table must be protected extremely well from unauthorized access. The internal key carries no intrinsic meaning and provides a user with no linking to a student's name, identifier, or other files without going through the confidential crosswalk table.

The Identifier Database and Locator System: A web page is recommended to provide timely access to a reference file of student identifiers. (The locator system might be named differently to avoid the impression that it locates students as opposed to locating identifiers. Some possible names are "Secure ID," "ID Fix," "ID Bank," and "ID." ID is used in the attachment as an acronym for Identifier Database.)

Two options exist for MDE's use of the locator.

a. The locator can serve simply as a reference for student information and the assignment and verification of student identifiers. The locator database would be updated each time a submission is received by MDE (possibly four times annually).

b. The locator can provide the added value of tracking the enrollments and withdrawals of students statewide. MDE would mandate that every time a student enrolls in a new school, even if the move was within the same district, that the registrar would connect to the locator and enter the entry date. This would provide an up-to-date reference for schools searching for information

as contrasted with Option a., which provides snapshots of student locations after each submission date.

What data elements should be included in the locator's database? Unfortunately, a student's name, birthdate, and gender do not reliably provide uniqueness to an individual. These three elements must all be included to provide a reasonable first query for matches.

A set of data elements describing a student's recent school enrollment adds significantly to the ability of a locator to find matches. These elements include: last school and district enrolled, school year or enrollment dates, and grade level enrolled.

A detailed set of personal data elements, although controversial, would be very helpful. These elements might include: mother's name, father's name, ethnicity, SSN, and siblings.

Value-Added Uses for the Identifier Database and the Locator System

In the consideration of the design and use of a locator, there are several valuable additional uses to which the information in the identifier database could be applied.

1. **Electronic Records Request:** The locator can offer a feature to send an electronic request for a student's record to the student's prior school. Transaction set 146, Request for a Student Record, ANSI X12 Standards, SPEEDE/ExPRESS, provides a format for such a request.

2. **Migrant Programs:** Students eligible for Title I Migrant program services could be identified. This process would be compliant with the Federal requirement for timely records exchanges for migratory families.

3. **Dropouts:** Tracking and verification of mobile students can reduce the reported dropout rates by providing a way to document transfers who otherwise would be considered dropouts.

4. **Placement:** The locator has the potential to provide a new school with valuable placement information about the student. Placement in proper courses, support services, and programs can save a new school the time and resources required for assessments. The student can be provided more continuous services and avoid changes that might be required upon completion of a reassessment or arrival of records from a prior school. In addition, students with special needs, e.g., vision or hearing modifications, emergency procedures, or free meals, can be accommodated. The inclusion of data elements useful for placement decisions changes the nature of the locator and raises extended confidentiality and access issues.

The drawback of including these value-added features in the initial locator is that each one raises additional issues and brings into the design process additional players and considerations. A prudent approach is to design the locator to accommodate these features and uses, but implement the primary objective of the locator—to verify and assign student identifiers.

Options for Districts

Districts may differ in their preference for use of a local management process for assigning from their block of identifiers versus use of a statewide locator for that purpose. Therefore, two options are presented below.

a. Districts receive from MDE a sufficiently large block of student identifiers (on a paper listing and/or in a file). Districts can distribute these to schools or establish a districtwide assignment process. Whatever their preference, the district must assure that an identifier is assigned only once, and that students with a previously assigned identifier are matched with their permanent identifier.

b. Districts may choose to require their schools to access the statewide identifier database (locator) for the assignment of identifiers. The locator would manage that process and provide the identifier for each new student.

Suggested Confidentiality for MDE Decisions Related to the Student Identifier

Public entities are not obligated to publish or provide to the public details related to systems and designs intended to protect the security and confidentiality of information systems. This protection is intended to prevent unauthorized access beyond that allowed by law or policy. Therefore, the details associated with Attachment B, and certainly any decisions made related to design and implementation of the student information system and the student identifier should be designated as internal documents, not for public distribution.

This paper was prepared for the Massachusetts Department of Education by Evaluation Software Publishing, Incorporated. The analyses and recommendations presented here are those of ESP. They may or may not reflect the thinking or any final decisions made by MDE.

ATTACHMENT A

Issues and Recommendations

Recommendation State Policy Decision Required	Social Security Number	State Identifier Local Pool	State Identifier State Pool	Local Identifier	Analysis and Recommendation
1. Uniqueness--At what level must the identifiers be unique?					An identifier must not be duplicated within a population or be an alias for a single individual within the population. Uniqueness must be maintained at the state level for a statewide identifier to ensure accuracy and completeness of data. The current identifier assigned by schools and districts to their students may not be unique across other districts. Statewide uniqueness can be achieved if districts assign identifiers that are unique within the district and begin with a district number. However, this complicates the process of verifying and using those numbers when students move across districts. There are certain benefits to using an identifier that is unique nationwide, e.g., SSN. SSN provides functionality for tracking former students into postsecondary education, verifying the identity of students across states, and exchanging useful data with other state and Federal agencies that provide services to families and individuals. These benefits can be achieved by collecting SSN irrespective of its use as the student identifier. RECOMMENDATION: Create student identifiers that are unique at the state level.
School	x	x	x	x	
District	x	x	x	x	
State	x	x	x		
Nation	x				
2. Burden--What level of burden should be imposed upon LEAs?					Burden is the time, effort, and resources required to implement the student identifier. This includes creating the system, assigning identifiers, verifying a identifiers, and entering identifiers where required. Burden includes effort to make the transition from an existing system to a new one. Burden must be limited to achieve compliance. Too high a burden will introduce unwanted errors. Burden must be balanced by benefit. Benefit must be assumed to be high because they are critical to the functionality of the entire student information system. The options that impose the least burden, school and district identifiers, fail to provide the functionality required. Burden is typically an issue to recognize and to manage rather than being one of the deciding issues when a student identification system is selected. RECOMMENDATION: Accept a moderate level of burden in exchange for the clear benefits from a functional student identifier within the information system.
None					
Minimal				x	
Moderate	x	x			
No Limit			x		

ATTACHMENT A

Issues and Recommendations

Recommendation State Policy Decision Required	Social Security Number	State Identifier Local Pool	State Identifier State Pool	Local Identifier	Analysis and Recommendation
3. Assignment–At what level will the identifiers be assigned?					This issue is related to burden. Assignment of the identifiers at the school level provides the quickest and least burdensome alternative. However, this issue is also related to uniqueness. The schools must follow a procedure that ensures unduplicated identifiers. Therefore, identifiers should be assigned at the lowest level possible without losing their uniqueness. A reality is that parents and students cannot be relied upon to carry their student identifiers from one school to the next. Mobile families too often cannot identify their last school/district, do not have records with them, and cannot remember student identifiers. RECOMMENDATION: Assign identifiers at the level where registration occurs. However, this does not preclude the use of state-assigned identifiers or a student locator system that provides the identifier to be assigned.
School		x		x	
District		x		x	
State			x		
Nation	x				
4. Timing–When will the identifier be assigned?					Schools need an identifier immediately upon enrollment of a new student. Certain forms are completed at that time and begin to go their separate ways. Ensuring that the student's identifier is on each form immediately saves changes and mismatches later. The difference between immediately and within a few hours is arguable. However, any system that takes days to assign an identifier presents a very different level of burden. RECOMMENDATION: The identifier should be assigned at registration and be available to schools before enrollment forms begin to be distributed to their respective offices.
Prior	x				
Immediate		x		x	
W/in Hours			x		
W/in Days			x		
At Reporting			x		
5. Verification Level–Where will the identifier be verified?					Verification ensures that the identifier is valid and correct. Valid means the identifier is included in the pool to be assigned. Correct means the identifier is accurately matched to the student. The identifier can be verified registration or later. Earlier verification means fewer changes will be required later. The closer the verification occurs to the parent and student, the higher the probability of accuracy. Verification at registration is best. The person registering the student must be authorized to assign the identifier. Verification conducted at the state level after submission of records is the least efficient. Correct identifiers can be challenged based upon duplications in the data elements used for verification. Verification is then delegated back to the school. State-level verification is required, but the frequency of potentially incorrect identifiers can be greatly reduced by adequate controls at registration. RECOMMENDATION: The identifier should be verified at the time of registration when parents and students are available to answer questions.
Local		x	x	x	
Other District				x	
State					

ATTACHMENT A

Issues and Recommendations

Recommendation State Policy Decision Required	Social Security Number	State Identifier Local Pool	State Identifier State Pool	Local Identifier	Analysis and Recommendation
6. Assignment, Verification Process—How will the identifier be assigned or verified?					Assignment includes 3 steps: First, the registrar determines if the student already has an identifier assigned. Second, the registrar secures the existing identifier or causes a new one to be assigned. Third, the registrar records the identifier in local records. The registrar can accomplish the first step by asking the parent or student, or records from a prior school. Parents may not have records, and may not precisely identify contacts for prior schools. Migrants' enrollment in a prior school may have been too brief to generate an official record. A statewide reference, such as a web page, can be queried to determine a previously assigned identifier. The second step could be accomplished using the reference to learn the identifier or assign a new one. Otherwise the registrar must contact the prior school. RECOMMENDATION: Registrars should assign and verify identifiers at the time of registration using a statewide reference for finding existing identifiers and a state-approved process for assigning new numbers
Query parent or student	x				
Contact prior school	x	x	x	x	
Examine transferred records	x	x	x	x	
Connect to a statewide database	x	x	x	x	
7. Verification Data Elements—What data elements are required for verification of a student's identifier?					When a student's identifier is in doubt, other unique combinations of information about the student must be used for verification. This is typically "directory information" as defined by FERPA along with a few other pieces of information included to increase the probability of describing a single student. The directory information could include name, birthdate, place of birth, gender, prior schools/districts of enrollment, enrollment dates, and grade level. Additional detail information could include parents' names, ethnicity, and SSN. Security and confidentiality issues must be considered. However, the more information available for query, the more likely existing identifiers will be found and used. RECOMMENDATION: As many of these data elements as practical should be included in the verification resource.
Directory information	x	x	x	x	
Detail personal information	x	x	x	x	
8. Confidentiality--Who may know the identifier?					FERPA, Massachusetts 603 CMR 23.00, and local policies will define the answer. If the identifier used is SSN, then SS laws would restrict access as well. If the identifier is a nominal code without intrinsic meaning, then it may be viewed as directory information. Prudent practice would call for the number to be treated as confidential, because knowledge of the number would place the holder one step closer to access to confidential information. RECOMMENDATION: Only education employees with a need to know should access the student identifier.
Public					
Local Ed Use	x	x	x	x	
State Ed Use Only	x	x	x	x	

ATTACHMENT A

Issues and Recommendations

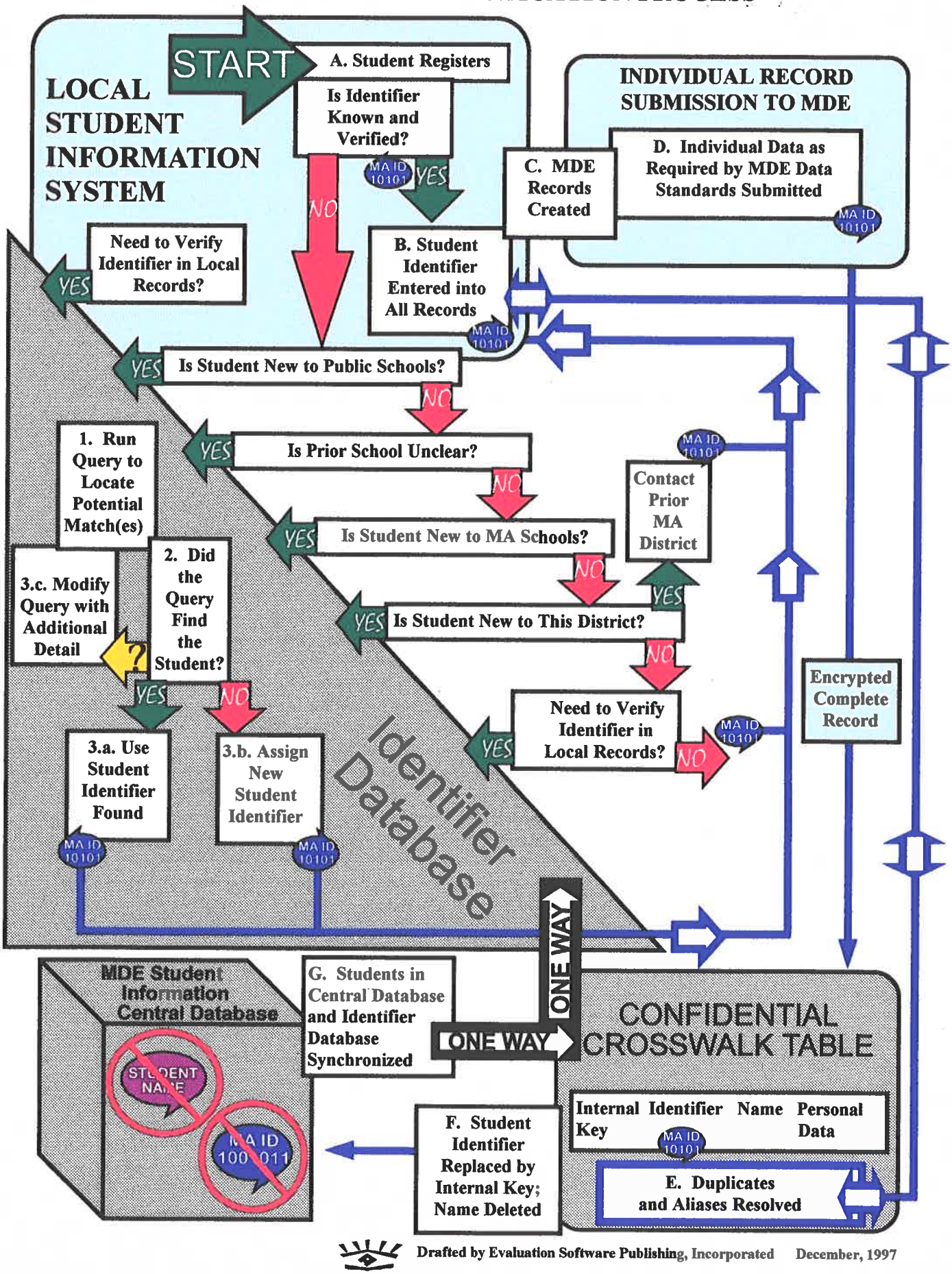
Recommendation / State Policy Decision Required	Social Security Number	State Identifier Local Pool	State Identifier State Pool	Local Identifier	Analysis and Recommendation
9. Imbedded Information--What meaning will be built into each number?					SSN has no important intrinsic meaning imbedded in the numbers. (Generally, the number may imply a region, year of assignment, or sequence, but any algorithm producing the numbers is obscure.) Imbedded information typically adds to the length of a number. For example, county/district codes add six characters to a number. Birthdate adds six or eight. A truly unduplicated, random number carrying no meaning has the advantage of requiring less restrictive security and confidentiality precautions. If the district number is imbedded, this could identify the student's first district of enrollment; however, that information can be carried in other fields within the database as well. RECOMMENDATION: There should be no imbedded information in the student identifier.
None	x	x	x	x	
Demographic		x	x		
Encrypted		x	x		
10. Length--How many characters can be in each number?					Shorter numbers can be entered, transcribed, and maintained with fewer errors. A common length for an identification number on scanner documents is 10. SSN is currently 9, 10 numbers have been discussed. To accommodate a million students without reusing numbers, requires a minimum of 7 numbers. Student mobility research estimates that half of student mobility is within the same district; half of the remaining half (or 25% of the 20% overall = 5%) is across state lines. Massachusetts may be higher with so many boarder states. New kindergarten or prekindergarten students could require about 10% new numbers annually. So with 5% for mobile students, 10% for kindergarten students, and a generous 5% for administrative margin, an annual use of 20% new numbers would require 4 million numbers across 15 years. To be prudent, adding additional digits provides a cushion for unanticipated events, i.e., the desire to imbed meaningful information. An 8- to 10-digit number appears to be a very safe minimum. RECOMMENDATION: Create a 10-digit student identifier.
<9		x	x		
9 to 10	x	x	x		
>10		x	x	x	

ATTACHMENT A

Issues and Recommendations

Recommendation State Policy Decision Required	Social Security Number	State Identifier Local Pool	State Identifier State Pool	Local Identifier	Analysis and Recommendation
11. Characters--What should be the nature of the characters?					Any number, letter, or symbol could be used. Symbols and letters present problems with recognition and accuracy in entering—especially when mixed with numbers. Certain letters (e.g., o, l, i, z, E, b/d, q/p) are sometimes confused. Using both numbers and letters provides for many more combinations and shorter identifiers. Problematic numbers and letters could be eliminated. Letters require too many bubbles on a scanner form. Letters come in capital and lower case forms that may or may not have meaning, but often cause confusion. Numbers are easier to distinguish, they can carry intrinsic and extrinsic meaning, and they are more "universal" across languages and cultures. Numbers can be assigned without risking the creation of meaningful and undesirable combinations as with letters. In the absence of considerations that require restricting the length of the identifier, use of only numerals is preferable. They are easy to distinguish. They can be entered with efficient keystrokes using a number pad. They require less space and are associated with less bubbling error on scanner forms. RECOMMENDATION: Only numerals should be used for the student identifier.
Numeric	x	x	x	x	
Numeric w/ Characters				x	
Alphanumeric				x	
12. Rubric--What conditions will be imposed on the numbers?					If an algorithm or imbedded meaning is to be used, then those considerations will answer this question. However, if a random number is used, then several rules can be followed to reduce data entry and clerical errors. Leading and final zeros are accidentally dropped. Consecutive identical numerals are entered too few or too many times. Eliminating these would reduce the available pool of numbers by about 15%. A final check digit is used to flag invalid numbers. This check digit could be the 10th character. If a check digit is used, and it is not allowed to be zero, then some of the numbers eliminated because they end in zero would be recovered. RECOMMENDATION: Use unduplicated random numbers with the exception of any with an initial zero, or any sequence of three or more identical numerals. Calculate a final nonzero check digit.
None	x			x	
No Lead Zero		x	x		
No Final Zero		x	x		
No Triple Numbers		x	x		

ATTACHMENT B: MDE STUDENT IDENTIFIER ASSIGNMENT AND VERIFICATION PROCESS



Drafted by Evaluation Software Publishing, Incorporated December, 1997