

GAIN Naming Conventions

Last updated: 06/18/2010

Authors:

**Melissa Ives,
Rod Funk,
Dionna Christian**

Table of Contents (click on a title to jump to that section)

[GAIN Instrument files](#)

[Item naming conventions](#)

[On the GAIN Instrument](#)

[Within SPSS datasets](#)

[Missing value codes](#)

[Scales and Indices Naming Conventions](#)

If you have questions or comments about the contents of this document, please contact the GAIN Analytic Team at GAINeval@chestnut.org

Chestnut Health Systems, Inc.

Lighthouse Institute

448 Wylie Drive

Normal, Illinois 61761

Internet: <http://www.chestnut.org>



Naming Conventions and Consistency Codes

To facilitate administration, scoring, interpretation and subsequent analysis, GAIN items are numbered and formatted the same regardless of version.

GAIN Instrument files

- GAIN pdf files begin with the abbreviation of the instrument (gi, gm, etc.), followed by an underscore and the abbreviated version number (5.2.1, 5.3.5, etc.), followed by another underscore and any special notation necessary for clarification.
 - Examples:
 - gi_5.6.3_csatsat_full.pdf: the GAIN-I version 5.6.3 Full with the GPRA module.
 - gi_5.6.2_core.pdf: the GAIN-I version 5.6.2 Core version used for sites that are not required to use the GPRA module as this version does not include the GPRA module.
 - gm_5.6.3_csatsat_core.pdf: the GAIN-M90 version 5.6.3 Core version with the GPRA module.
 - The first two portions of the version number indicate the instrument version (05.06 is the 5.6 version), while the third digit indicates paper-only changes (such as correcting typos or adding clarifying statements; 05.06.02 indicates this kind of revision compared to 05.06.01).
 - Beginning with version 5.6, the instrument consists of either the full or core and either regular or CSAT. There is no longer a distinction for regional as this is now incorporated as the regular GAIN without the GPRA module.
 - The information specifying version (instrument abbreviation, version number, and special notation) can be found in the bottom left hand corner of the paper version of the instrument.
 - GAIN exported data files for Legacy ABS (for versions through 5.5) begin with the full name of the instrument (e.g. GAIN-I, GAIN-M90, etc.) followed by what type of file it is. Versions 5.6 and higher use GAIN ABS and include “GAIN ABS” in the name of each exported file.
 - Examples:
 - GAIN-I.txt: The GAIN-I data file to read into SPSS from Legacy ABS
 - GAIN-Q_Notes.rtf: The GAIN-Q notes file from Legacy ABS
 - GAINABS-M90.txt: The GAIN-M90 data file to read into SPSS from GAIN ABS
 - GAINABS-QM_SPC.txt: The GAIN-Q specification file for reading data into SPSS from GAIN ABS

Item naming conventions (GAIN-I, GAIN-M90)

GAIN items consist of a combination of letters and numbers: B2, B4j, S2s1a, etc.

- The first letter refers to each section of the instrument:
 - **A** Administration
 - **B** Background and Treatment Arrangements

- Substance Use
 - Physical Health
 - Risk Behaviors and Disease Prevention
 - Mental and Emotional Health
 - Environmental and Living Situation
 - Legal
 - Vocational
 - Z (end)
 - X (across dataset ID information)
 - CG for CSAT GPRA items (CSAT versions only)
 - XADM for administrative items completed by staff at the end of the assessment
 - XDSM for diagnostic items (versions through 5.5)
 - XSS for Special Study items (versions through 5.5)
- The next digit refers to the question number within a section: S1, S2, S3. Letters, usually indicating a series of interrelated questions, follow: S2a, S2b, S2c. Occasionally a number follows, in which case the second digit is separated from the third by an underscore: L7_1, L7_2, L7_3 (which is done to avoid, for example, confusing question 3, subresponse 1 with question 31).
 - Any subsequent digits alternate between numbers and letters in outline format.
 - If there is an instruction to “Please describe” and a line starting with a *v* after it, then in addition to the main variable there is a text variable with the same name plus a *v* (e.g., S7b99 and S7b99v). Note that the *v* also means that any text written there should be “verbatim” (entered as closely to the participant’s own words as possible).
 - The letters *i*, *o*, and *l* are always skipped in the item numbering. This was done intentionally to avoid confusion in documentation and analysis with the numbers 0 and 1.
 - We have included an “other” in most groups of questions and given it a value of 99 so that other questions could be added to the group for a special population or study.
 - Dates should be recorded in the mm/dd/yyyy format, e.g., 03/06/1996 for March 6, 1996.

Within SPSS datasets

- **d** suffix or ‘**any**’ prefix indicates a dichotomous version of a variable
- **g** suffix indicates a trichotomy recoded from a scale or index based on the low, moderate and high cut points.
- **c** suffix indicates a recoded (not dichotomous) version of a variable into more than 3 groups.
- **l** suffix indicates a lifetime variable
- **y** suffix indicates a past-year variable
- **m** suffix indicates a past-month variable
- **m** prefix indicates a missing but replaced version of a variable
- **p** suffix indicates a proportional version of a variable (divided by range), the result of which ranges from 0 to 1

- **r** prefix indicates an item has been reversed (low values converted to high and vice versa)
- **t** (or **T**) prefix indicates a T-score version of a variable
- **vld** suffix indicates a count of valid items in a scale or index
- **v** suffix generally indicates a verbatim field (occasionally, a GAIN item that is not a verbatim ends in *v*; e.g., item S9v).
- **z** prefix indicates a z-score version of a variable
- **_#** suffix, generally found in the horizontal file, indicates an item's wave (xobs) of origin; e.g., xobsdt_3 is the date of the 3-months interview
- **_0** is a special case of the **_#** suffix in a vertical file. In this case, it indicates a baseline variable that has been carried down to other waves (e.g., XOBSDT_0 Date of intake interview or B2a_0 Age at intake) for use in particular analyses such as Mixed linear model analysis.
- **_1** suffix indicates the item is from the last wave of available data
- **np_1** suffix indicates NOMS outcomes of no problems at the last wave (intake version is **np_0**)
- **90** suffix indicates that an item is "in the past 90 days" (absence of 90 is not necessarily the same as "not in the past 90 days")

Missing value codes (all are denoted with a minus sign before the digit)

Data may be missing for a variety of reasons. When possible, missing data is assigned a code to identify why it is missing, called a consistency code. All consistency codes are denoted with a minus sign before the digit.

- **-3:** Site chose in advance to not ask this item.
- **-4:** (version 5.6 and higher) Item accidentally omitted, invalid response, missing data.
- **-6:** Data omitted from exported dataset because of confidentiality issues.
- **-7:** Client refused to respond.
- **-8:** Item accidentally omitted, invalid response or client responded DK (versions through 5.5); Client responded DK (version 5.6 and higher).
- **-9:** Legitimately skipped item (replaced in dataset with logical value, usually zero).

Scales and Indices Naming Conventions

Scales are classical scales made up of internally consistent indicators that are combined to estimate an underlying latent trait. Summative indices consist of a linear combination of different types of observed indicators. These indicators (items) need not be internally consistent.

- Scales and indices are named so that higher values indicate more of the scale. For example, on the Training Activity Scale, higher values mean more training activity; on the Cognitive Impairment Screen, higher values mean greater impairment.
- Scale and index variable names are abbreviated using the first letters of the words in the scale name: e.g., the Substance Problem Scale–Lifetime is SPSL and the Parent Activity Index is called PAI.
- Sometimes, to avoid having two variables with the same name, additional letters are used: e.g., SxRS is the Sexual Risk Scale.

- A number at the end of a scale name indicates the number of items used in creating that variable. It is usually used to indicate that the number of items has changed across versions: for instance, GMDS26 is based on 26 items in version 5 of the GAIN, while in older versions, this scale had only 21 items and was named GMDS21.
- Scales or indices ending with a lowercase *p* were calculated using the proportional versions of items. Proportional versions are the item value divided by the item range. (days of use in the past 90 days/90). These scales range from 0 to 1.