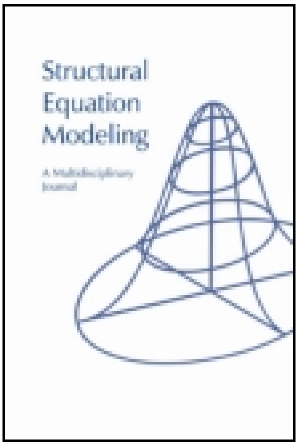

Structural Equation Modeling: A Multidisciplinary Journal
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/hsem20

Conducting Confirmatory Latent Class Analysis Using M*plus*

W. Holmes Finch [a] & Kendall Cotton Bronk [a]
[a] Ball State University

Available online: 07 Jan 2011

Structural Equation Modeling, 18:132–151, 2011

ISSN: 1070-5511 print/1532-8007 online
DOI: 10.1080/10705511.2011.532732

Psychology Press
Taylor & Francis Group

Conducting Confirmatory Latent Class Analysis Using M*plus*

W. Holmes Finch and Kendall Cotton Bronk
Ball State University

Latent class analysis (LCA) is an increasingly popular tool that researchers can use to identify latent groups in the population underlying a sample of responses to categorical observed variables. LCA is most commonly used in an exploratory fashion whereby no parameters are specified a priori. Although this exploratory approach is reasonable when very little prior research has been conducted in the area under study, it can be very limiting when much is already known about the variables and population. Confirmatory latent class analysis (CLCA) provides researchers with a tool for modeling and testing specific hypotheses about response patterns in the observed variables. CLCA is based on placing specific constraints on the parameters to reflect these hypotheses. The popular and easy-to-use latent variable modeling software package M*plus* can be used to conduct a variety of CLCA types using these parameter constraints. This article focuses on the basic principles underlying the use of CLCA, and the M*plus* programming code necessary for carrying it out.

Latent class analysis (LCA) is an increasingly popular analytic technique useful for identifying latent groups based on a set of observed response variables, which can be either dichotomous or polytomous. It is important to note here that a variant of LCA known as latent profile analysis can be used when the observed variables are continuous, but the focus of this article is on LCA with dichotomous observed variables. Table 1 includes a simple taxonomy for organizing the appropriate analysis by the type of research question to be addressed and the type of data available. These are merely examples of the many research questions that can be addressed by these models, and are not intended to be an exhaustive list. Typically, LCA is carried out in an exploratory manner where there does not exist a strong a priori hypothesis regarding the number or nature of the latent classes underlying the data (Hoijtink, 2001). In such cases, a researcher can fit several proposed models to the data with each differentiated by the number of latent classes, and compare the resulting fit indexes to determine which best corresponds to the observed data.

This exploratory analysis approach works under the implicit presumption that there is not a well-developed theory regarding the nature of latent groups to be found in the population (Laudy, Boom, & Hoijtink, 2005). However, in cases where substantive theories regarding

Correspondence should be addressed to W. Holmes Finch, Department of Educational Psychology, Ball State University, TC 521, Muncie, IN 47306, USA. E-mail: whfinch@bsu.edu

TABLE 1
Taxonomy of Models for Latent Categorical Variables

Type of Research Question	*Type of Observed Variable*	
	Categorical	*Continuous*
Exploratory	Latent class analysis (How many latent classes underlie a set of categorical observed variables?)	Latent profile analysis Cluster analysis (How many latent classes underlie a set of continuous observed variables?)
Confirmatory	Confirmatory latent class analysis (Are there three latent classes underlying a set of categorical variables, with Group 1 having higher response probabilities than Group 2 and Group 3 having the lowest probabilities, as theory would suggest?)	Confirmatory latent profile analysis (Are there three latent classes underlying a set of observed continuous variables such that Group 1 has the highest mean values, followed by Group 2, which in turn has higher means than Group 3, as theory would suggest?)

Note. Example research questions associated with each analysis are shown in parentheses.

the number and nature of these latent classes have been developed, exploratory LCA might be inefficient, not taking advantage of this prior knowledge. Confirmatory LCA (CLCA) is an alternative approach to latent class modeling that allows for the formulation of specific hypotheses regarding the nature and number of latent classes in the data. These hypotheses are expressed as a set of parameter constraints for an estimated LCA model (Croon, 1990). The goal of this article is to demonstrate how such parameter constraints can be used in a common latent variable modeling software package, M*plus,* to carry out CLCA. First, we briefly introduce the basic LCA model, and then describe how constraining parameter values can be used to express specific hypotheses regarding latent classes in the population. We will then present several examples of CLCA using a set of dichotomous items taken from a survey on adolescent purpose.

LATENT CLASS ANALYSIS

The basic LCA model is described in some detail by McCutcheon (2002). Assume that data have been collected for four observed, dichotomous variables, $X1$, $X2$, $X3$, and $X4$, and that there exists a latent categorical variable Y, which accounts for the relationships among these four observed variables. The LCA model linking the latent and observed variables can then be expressed as:

$$\pi_{ijklt}^{X1X2X3X4Y} = \pi_t^Y \pi_{it}^{X1|Y} \pi_{jt}^{X2|Y} \pi_{kt}^{X3|Y} \pi_{lt}^{X4|Y} \quad (1)$$

where

π_t^Y = Probability that a randomly selected individual will be in latent class t of latent variable Y

$\pi_{it}^{X1|Y}$ = Probability that a member of latent class t will provide a response of i for observed variable $X1$

$\pi_{jt}^{X2|Y}$ = Probability that a member of latent class t will provide a response of j for observed variable $X2$
$\pi_{kt}^{X3|Y}$ = Probability that a member of latent class t will provide a response of k for observed variable $X3$
$\pi_{lt}^{X4|Y}$ = Probability that a member of latent class t will provide a response of l for observed variable $X4$

The LCA model in Equation 1 asserts that the observed variables are conditionally independent given a particular class in Y (Goodman, 2002). This notion of conditional independence is very similar to local independence in the context of item response theory, which states that when the latent trait influencing responses to items on an instrument is held constant, individuals' responses to any two items are statistically independent. As an example, take an individual from the population who has the following probability values for the three classes in Y: $\pi_1^Y = 0.6$, $\pi_2^Y = 0.25$, and $\pi_3^Y = 0.15$. These results indicate that the individual is most likely to be in Class 1 of the latent variable, with only a 1/4 chance of being in Class 2 and a less than 1/5 chance of being in Class 3. In addition, assume that observed variable $X1$ is a survey item asking whether an individual hopes to pursue a career helping other people after finishing college. A value for $\pi_{Yes\_1}^{X1|Y}$ of 0.75 would indicate that an individual in the first class of the latent variable would have a fairly high likelihood of responding "Yes" to this item. Another way to interpret this result would be that most individuals in latent Class 1 plan to help others after finishing college. The degrees of freedom for the latent class model with four indicators are calculated as $DF = (IJKL - 1) - [(I + J + K + L - 4)T - 1]$. Here, I, J, K, and L represent the number of categories in each of the observed response variables, and T is the number of latent categories.

ASSESSMENT OF FIT FOR LATENT CLASS MODELS

LCA involves the estimation of two types of parameters: (a) the probability of a particular response for an observed variable conditional on latent class membership, and (b) the probability of being in a specific latent class, t. Estimation of these parameters can be carried out using maximum likelihood estimation (MLE) via the EM algorithm, as is done in the M*plus* software package (B. O. Muthén, 2001), and model fit can be assessed using a variety of statistical tools (Nylund, Asparouhov, & Muthén, 2007). Nylund et al. (2007) conducted an extensive simulation study comparing a large number of these tools and found that among the information criteria, the sample size adjusted Bayesian information criterion (aBIC) was superior to alternatives such as the Akaike information criterion (AIC), the consistent AIC, and the standard BIC. The aBIC takes the likelihood ratio statistic and applies a penalty for an increased number of model parameters. It is calculated as:

$$aBIC = \chi^2 - df * [\ln(N^*)] \tag{2}$$

where

df = model degrees of freedom
$N^* = \left(\frac{N+2}{24}\right)$

The aBIC is used for comparing the fit of multiple models, with lower values indicating relatively better model fit.

In addition to the information criterion, Nylund et al. (2007) also examined the performance of three hypothesis testing approaches to assessing model fit: the chi-square-based likelihood ratio test (LRT), the Lo–Mendell–Rubin (LMR) test, and the bootstrap likelihood ratio test (BLRT). The LRT is not appropriate for comparing mixture models with differing numbers of classes because it does not follow the chi-square distribution under the null hypothesis of no difference in model fit when testing the number of classes (McLachlan & Peel, 2000). In contrast, the LMR statistic is appropriate for comparing mixture models with differing numbers of classes because it does not rely on the chi-square distribution for the difference in model likelihood values, instead using an approximation of this distribution to obtain the appropriate p values. A significant LMR result indicates that the mixture model with k classes fits the data better than the simpler $k-1$ class model. The BLRT also allows for the comparison of likelihood values for mixtures with differing numbers of classes by resampling from the null hypothesis of no difference. A complete discussion of the BLRT appears in McLachlan and Peel (2000). It should be noted that whereas these tests only allow for comparisons of models with differing numbers of latent classes and the same parameterization, the aBIC statistic can be used to compare models with the same number of latent classes, different parameterizations, or both. As noted earlier, the aBIC is based on the log-likelihood value, which provides information about how well a given model fits the observed data. Therefore, it is possible to compare any two models using the aBIC whether they have different numbers of parameters, unlike with the LMR and BLRT tests.

CONFIRMATORY LATENT CLASS ANALYSIS

As mentioned previously, most applications of LCA in practice involve exploratory analyses in which no a priori hypotheses regarding the nature of latent classes are explicitly tested (Laudy et al., 2005). In such cases, researchers do not attempt to explicitly test any theories about underlying groups in their substantive area, but rather allow the data to suggest the number and nature of such groups. However, in many fields prior work might provide the researcher with ideas regarding the characteristics of latent groups underlying the data, which can be explicitly tested. To develop and test such models, a set of parameter restrictions must be used to express these hypotheses explicitly. McCutcheon (2002) describes three types of parameter constraints that can be used in CLCA modeling: (a) equality restrictions, (b) deterministic restrictions, and (c) inequality restrictions.

In the case of equality restrictions, a researcher might wish to test that one or more item parameter values are equal across latent classes. For example, in his description of using CLCA with items measuring antisocial behavior, B. O. Muthén (2001) used an example in which two latent classes were constrained to have equal likelihoods for having broken into a building. Using such restrictions, Muthén was able to explicitly express theories regarding the expected subtypes of antisocial behavior in terms of expected common and distinct response patterns by individuals in the sample.

Deterministic model restrictions focus on testing whether conditional response probabilities equal some specific value, often 0 or 1, for one or more latent classes (McCutcheon, 2002).

For example, suppose a researcher hypothesizes that the population contains a latent class that does not exhibit antisocial behavior. He or she might believe that no members of this class will endorse an item stating that they intend to injure another person, which can be expressed in a CLCA by restricting the conditional probability of endorsing this item a priori to be 0 for one of the latent classes.

A third type of CLCA parameter constraints involves using inequality restrictions to test hypotheses regarding the relative likelihood of latent classes endorsing an item. For example, the researcher interested in antisocial behavior might believe that latent classes in the population can be ordered based on their likelihood of endorsing the item "seriously threaten another individual" in the following way: $\pi_{11}^{X1|Y} > x_{12}^{X1|Y} > \pi_{13}^{X1|Y} > \pi_{14}^{X1|Y}$. Such a set of constraints can be explicitly modeled in the CLCA analysis to determine if this pattern actually exists in the population.

It is presumed that the hypotheses tested with CLCA come from prior research, clinical observation, or both, much in the way that hypotheses assessed using confirmatory factor analysis come from prior knowledge in the substantive area being studied. Each of these three approaches to setting parameter constraints can be carried out using the popular M*plus* software package (L. K. Muthén & Muthén, 2008), with the Mixture model add-on option. In this article, we describe how CLCA using parameter restrictions can be conducted in M*plus* and then provide an extensive example demonstrating each of these.

PARAMETER CONSTRAINTS IN M*PLUS*

As described previously, the conduct of CLCA involves the placement of constraints on model parameters (typically conditional probabilities) that reflect the substantive hypotheses proposed by the researcher. In M*plus* these constraints are expressed using variable threshold values, which for dichotomous variables such as those used in this article, are rescaled probabilities for a particular response category. The relationship between the response probability (P) and the variable threshold (τ) takes the form (B. O. Muthén, 2001):

$$P = \frac{1}{1 + e^{-\tau}} \tag{3}$$

Thus large positive thresholds indicate the probability of a specific response value is relatively low, whereas large negative values suggest that the probability of the response is relatively high. In the *M*plus *User's Guide for Version 5*, L. K. Muthén and Muthén (2008) provide some guidelines for interpreting and using thresholds, suggesting that a value of $+3$ represents a very low probability of a particular variable response, whereas a -3 reflects a very high probability. Indeed, in the context of an item response a $+3$ threshold translates to a probability of endorsing the item of 0.047, whereas a -3 translates to a probability of item endorsement of 0.953. In addition, they suggest that $+1$ and -1 threshold values can be interpreted as low and high probabilities, respectively. Once parameter estimation is completed, a determination must be made regarding the fit of the model to the data using the fit statistics described previously and recommended by McCutcheon (2002). The following is a brief description of the data that are used in the following examples.

Purpose in Life

Recently the youth development literature has experienced a sea change. Previously researchers focused on addressing young people's shortcomings and weaknesses, whereas today more attention is paid to enhancing youths' talents and capacities. Identifying developmental problems can be a relatively straightforward task, but identifying signs of optimal youth development can present a challenge. One of the key guideposts researchers have recently begun to point to as an important indicator of positive youth development is the presence of an inspiring and prosocial purpose in life (Benson, 2006; Damon, 2009). A purpose in life is a stable and generalized intention to accomplish something that is simultaneously meaningful to the self and leads to productive engagement with the world beyond the self (Damon, Menon, & Bronk, 2003). There are two components of this definition distinguishing purpose from the broader context of meaning in life: (a) a purpose can be viewed as a long-term goal, and (b) a purpose is personally meaningful, but it also has a prosocial desire to have an impact on the world beyond the self.

Participants

The data used in the following examples included 153 adolescents and 237 emerging adults who either lived in or attended college in the Midwest ($N = 390$). The sample was 47% male and predominantly White, representing the ethnic makeup of the Midwest data collection location.

Measures

Participants completed the Revised Youth Purpose Survey (Bundick et al., 2006), which was created by members of the Stanford Center on Adolescence to assess the prevalence and types of purpose present among adolescents. Purpose is assumed to consist of a subset of sources of meaning. Although individuals can find meaning in either externally directed aims (e.g., helping those who are less fortunate) or internally directed pursuits (e.g., seeking fame and fortune) purposes only include those sources of meaning that include a desire to have an impact on the broader world (e.g., the intention to work toward a cure for cancer). Therefore, participants were asked to rate both internally oriented concerns and externally directed pursuits. The list of types of purpose intentionally included more externally directed aims because the authors of the survey were primarily interested in discovering more about the purpose construct. The specific types of purpose included in this study were drawn in part from studies of young people's sources of meaning conducted by De Vogler and Ebersole (1980, 1981, 1983) and Showalter and Wagener (2000), and adapted by the Stanford Center on Adolescence youth purpose research team. The types of purpose are listed in Table 2.

PARAMETER ESTIMATION FOR CLCA MODELS

Four of the purpose items were used to assess a series of hypotheses regarding the nature of purpose among adolescents particularly as it pertains to their need to be creative and change the way people think versus their desire to have fun and make money. These items, which

TABLE 2
Types of Purpose

Internally Directed Aims	*Externally Directed Aims*
Live life to the fullest	Help others
Make money (y10)[a]	Serve God or a higher power
Have fun (y15)[a]	Make the world a better place
Be successful	Change the way people think (y4)[a]
Have a good career	Create something new (y5)[a]
	Make things more beautiful
	Fulfill my obligations (to others)
	Do the right thing
	Discover new things about the world
	Earn the respect of others
	Serve my country
	Support my family and friends

[a]Variables included in the confirmatory latent class analyses with (variable name).

appear in Table 2 along with their M*plus* variable names, were converted from a 7-point Likert scale ranging from *strongly disagree* to *strongly agree* to a binary scale (1 = *agree,* 0 = *do not agree*). This change to the data was made for two reasons. First, the researchers participating in this study believed that youth would generally either endorse or not endorse the types of purpose being asked about on the instrument. In other words, although the items were originally placed on a 7-point Likert scale, subsequent work has led researchers to believe that most youth actually think of these types of purpose in a yes–no way. In addition, because response patterns on the 7-point scale were indeed bimodal for these items, with the vast majority of respondents tending to either agree or strongly agree or disagree or strongly disagree with the statements, this supposition appears to be upheld empirically. Therefore, the decision was made to rescale the data to conform to the latest thought in the field that was also buttressed by empirical evidence. However, it should be noted that making such changes to the data is not without consequence. First of all, the participants did provide responses based on a 7-point Likert scale, even though the vast majority was at one end or the other. Therefore, the psychometric properties of the items are no longer known because reliability and validity analyses done previously would only apply to the full 7-point scale. Second, because some of the respondents did have scores that were in the middle of the scale, collapsing categories does result in a loss of information. In this instance, however, it was determined that because the bimodal data matched the dichotomous distribution of responses that researchers expected, this combining of categories was reasonable.

Based on prior research in the area of purpose, it is hypothesized that there exist four latent classes with regard to creativity and personal gain: (a) those who want to change the way people think, create something new, and have fun, but are relatively unconcerned about making money; (b) those who only want to have fun and are unconcerned about changing the way people think, creating something new, or making money; (c) those who are unconcerned about changing the way people think or creating something new but who want to make money and have fun; and (d) those who would like to make money, have fun, change the way people think, and create something new. In the following sections, we provide examples of CLCA to investigate the proposed latent classes using the parameter constraints discussed earlier.

EQUALITY CONSTRAINTS

One type of hypothesis for the purpose data is that certain of the four latent classes share common response probabilities on the four items. An example of a set of proposed response parameter constraints for this model appears in Table 3. The presence of a common number for two or more classes on a given item indicates that the groups are constrained to have a common threshold value for that item. Conversely, classes with different numbers for a given item are allowed to have different threshold values and therefore different probabilities for endorsing the item. As an example consider Classes 1 and 2, which are hypothesized to have a common threshold parameter on the items “make money” and “have fun,” but not on “change the way people think” or “create something new.” Note that in keeping with the hypothesis briefly described earlier, all four classes are expected to have a common threshold value on the item “have fun.” The M*plus* commands for conducting this analysis appear in the Appendix. The full set of commands as well as output that are presented in this article can be obtained by contacting the first author.

Only the four variables of interest are used for this analysis, although all 17 items and the student identifier are read in. The CLASSES statement defines the latent class variable as being named “c” and having four classes. The ANALYSIS command indicates that we are conducting a MIXTURE model analysis, and the STARTS subcommand tells M*plus* the number of random sets of starting values and the number of optimizations to use in the final estimation of parameter values. The default is 10 random sets of starting values and two optimizations. However, it is recommended that when more than two latent classes are present, more random starts be used to avoid arriving at local maxima for parameter estimates (L. K. Muthén & Muthén, 2008).

It is in the MODEL command where the parameter constraints displayed in Table 3 are made explicit. The item thresholds are defined separately for each latent class, such as for item y4, **[y4$1*-2] (1);**. The item name (**y4**) is given, followed by **$1** indicating the first (and only for this dichotomous item) threshold value. The ***-2** provides a starting value for the estimation of the threshold value (a starting probability of 0.881) and was selected because it fell between the high and very high probability guidelines in the M*plus* manual, corresponding to the expectation that this group is very likely to endorse the item, although perhaps not at rates exceeding 0.95. Note that for latent Classes 2 and 3, the starting value for this item was 2, corresponding to a probability of 0.119. The **(1)** numbers this parameter value and serves as the method by which equality constraints are made, as described previously. The numbering scheme for parameter constraints displayed in Table 3 is used here. In this way, we can ascertain

TABLE 3
Hypothesized Response Patterns for the Four-Class Confirmatory Latent Class Analysis Model of Future Purpose

Item	*Class 1*	*Class 2*	*Class 3*	*Class 4*
Change the way people think (y4)	1	2	2	1
Create something new (y5)	3	4	4	3
Make money (y10)	5	5	6	6
Have fun (y15)	7	7	7	7

the degree to which the hypothesized pattern matches the actual data. A brief discussion of a very similar CLCA problem using such constraints appears in B. O. Muthén (2001).

Although in this example we provide starting values for the threshold parameters, but do not constrain them to be a particular value, it is possible to constrain the threshold to be a specific value for one or more groups. The decision on whether to allow the threshold to be estimated (as in this example) or to be set to a predetermined value (appearing later) is based on the goals of the researcher and the presence (or not) of hypotheses for the parameter values. In this case, the researcher does not have an a priori hypothesis about the specific value of the threshold in the population, although he or she believes it will be low, and therefore allows the parameter to be estimated freely. As described later, the researcher can also force a threshold to be a specific value.

We can be comfortable that parameter estimation converged normally due to the lack of a message warning us about convergence problems. Had there been such a convergence problem, a warning message would have been generated, and the results contained in the output could not be relied on to be accurate. Note that we did receive a warning indicating that when estimating models with more than two latent classes, we should increase the number of random starts to avoid the problem of maximum likelihood converging to local maxima. We have done this with the STARTS command, as discussed earlier. The fact that each of the 10 log-likelihoods reached the same final value is an indication that the algorithm did not converge to local maxima for any of the 10 tries, but rather converged to a single (presumably global) value. On the other hand, if several of the log-likelihood values differed from one another, this would suggest that some of the random starts had resulted in convergence to local maxima.

RANDOM STARTS RESULTS RANKED FROM THE BEST TO THE WORST LOGLIKELIHOOD VALUES

Final stage loglikelihood values at local maxima, seeds, and initial stage start numbers:
-853.769 347515 24
-853.769 749453 33
-853.769 284109 82
-853.769 352277 42
-853.769 761633 50
-853.769 391179 78
-853.769 626891 32
-853.769 314084 81
-853.769 685657 69
-853.769 533738 11

WARNING: WHEN ESTIMATING A MODEL WITH MORE THAN TWO CLASSES, IT MAY BE NECESSARY TO INCREASE THE NUMBER OF RANDOM STARTS USING THE STARTS OPTION TO AVOID LOCAL MAXIMA.

THE MODEL ESTIMATION TERMINATED NORMALLY

The fit statistics for the four class model appear in Table 4. The aBIC value for this model was 1715.616, and both the LMR and BLRT tests were statistically significant ($p < .05$). The

TABLE 4
Fit Statistics for Competing Models

Model	*LMR p Value*	*BLRT p Value*	*aBIC*
Four-class	<.0001	<.0001	1,715.616
Three-class	.0986	.0810	1,732.302
Two-class	<.0001	<.0001	1,731.880
Deterministic	<.0001	<.0001	1,736.430
Inequality 1	NA	<.0001	1,729.128
Inequality 2	NA	<.0001	1,714.097
Inequality 2b	NA	<.0001	1,715.967

Note. LMR = Lo–Mendell–Rubin; BLRT = bootstrap likelihood ratio test; aBIC = adjusted Bayesian information criterion.

aBIC value can be used to compare the fit of this model with that of others. The LMR and BLRT tests are comparing the fit of four latent classes versus that of three, and in this case indicate that the four-class solution provides the better fit.

The threshold values and proportion of individuals endorsing the items for the four-class model as well as the latent class sizes appear in Table 5. Latent Classes 1 and 4 were the largest, with more than 100 participants in each. When interpreting thresholds, it is important to remember that large positive values indicate a lower likelihood of individuals endorsing the item, whereas large negative values suggest just the opposite. For this example, the very large estimates of 15.000 for some of the items mean that for the latent class in question, the likelihood of endorsing these items is extremely small. Indeed, in each of these cases the proportion of individuals doing so was 0. On the other hand, the threshold estimate for the item "have fun" was −3.656 for all four classes, which translated into 97.5% of each group endorsing this item. Latent Classes 1 and 4 both had slightly negative threshold values for the first two items, indicating that they were more likely to endorse these than were members of the other two latent classes, and more than 65% of the members in each group did so.

TABLE 5
Threshold Parameter Estimates and Proportion Endorsing Items for the Four-Class Confirmatory Latent Class Analysis Model of Future Purpose

Item	*Class 1 (120)*	*Class 2 (30)*	*Class 3 (48)*	*Class 4 (176)*
Change the way people think (y4)	−0.808	15.000	15.000	−0.808
Create something new (y5)	−0.750	15.000	15.000	−0.750
Make money (y10)	15.000	15.000	−3.604	−3.604
Have fun (y15)	−3.656	−3.656	−3.656	−3.656
Proportion in Each Group Endorsing Item				
Change the way people think (y4)	.692	0	0	.692
Create something new (y5)	.679	0	0	.679
Make money (y10)	0	0	.974	.974
Have fun (y15)	.975	.975	.975	.975

Based on these results, the proposed four latent class structure described earlier seems plausible, although alternatives to this model are discussed later. Given the response patterns contained in Table 5, it is possible to characterize these latent classes. Class 1 is made up of individuals who plan on taking a creative role in society and changing the way people think, but who are not particularly concerned about whether this role will result in a high income. In contrast, Class 3 consists of those who are primarily concerned with making money and having fun, but who have little interest in changing the way people think or in creating something new. Class 4 is made up of people who are looking forward to both being creative/changing the way people think and making money/having fun. Finally, Class 2 appears to contain those participants who only want to have fun and have little or no interest in either being creative, changing the way people think, or making money.

Although the latent classes already described do appear to correspond with those that were originally hypothesized in Table 3, two alternative equality constraint models were also considered here. It is important to note that the exploration of this and other alternative models in this article is designed to be primarily pedagogical in nature. In actual practice, a researcher using CLCA would base his or her decisions regarding the models to test on hypotheses drawn from literature in the area of interest. Using the constraints common in CLCA for exploratory analyses in an atheoretical manner is not recommended because it creates the possibility of making substantive conclusions based on sampling variation rather than theoretically supportable empirical findings. It is key that all analyses be guided by substantive hypotheses, although researchers need not be limited to only a single one of these. In much the same manner that those using structural equation modeling might have competing hypotheses, based in theory, that can be compared with one another, so can the researcher using CLCA have competing hypotheses about the nature of group membership in the population. The first of these asserts that only three classes actually exist in the population, corresponding to Classes 1, 3, and 4 in Table 3. The model commands in M*plus* for estimating this alternative model appear in the Appendix. Note that only the CLASSES command was changed to reflect the presence of three rather than four latent classes: **CLASSES = c (3);**

The fit statistics for this model appear in Table 4. We can determine that the four-class model fit better than the three-class model based on a comparison of aBIC, for which the four-class model had a lower value. In addition, the LMR and BLRT tests for the four-class model indicated that it fit the data better than the three-class model. The results of LMR and BLRT corresponding to the three-class model in Table 4 compare the fit of this model with a two-class model. The results are not statistically significant for either test ($p = .0986$ and $p = .0810$, respectively), indicating that the three-class model does not fit the data significantly better than the two-class model.

In addition to the relatively poor fit indexes, the pattern of thresholds and corresponding proportion of individuals endorsing the items for the three classes suggest that this solution was not optimal (see Table 6). The a priori hypothesis was that the classes would consist of individuals who were likely to endorse all of the items, those who would endorse all items except for “make money,” and those who would only endorse the items “make money” and “have fun.” However, the results presented in Table 5 reveal that the group endorsing all items except for “make money” did not emerge. Rather, there were two classes whose members were likely to endorse all of the items, albeit Class 2 had a somewhat lower probability of doing so. This result would suggest that the three-class hypothesis does not seem plausible.

TABLE 6
Threshold Parameter Estimates and Proportion Endorsing Items for the Three-Class Confirmatory Latent Class Analysis Model of Future Purpose

Item	*Class 1 (103)*	*Class 2 (52)*	*Class 3 (219)*
Change the way people think (y4)	−1.952	−1.952	0.268
Create something new (y5)	−15.000	−15.000	0.478
Make money (y10)	−2.354	−0.160	−0.160
Have fun (y15)	−2.335	−2.335	−2.335
Proportion in Each Group Endorsing Item			
Change the way people think (y4)	.876	.876	.433
Create something new (y5)	1.000	1.000	.383
Make money (y10)	.913	.540	.540
Have fun (y15)	.912	.912	.912

Finally, a two latent class solution was considered, in which one class was characterized by low threshold values (high probability of endorsement) on all four items, and the other was characterized by low thresholds on the items "make money" and "have fun" and high thresholds on "change the way people think" and "make something new." The fit statistics for this model appear in Table 4. The aBIC value suggests that the two-class model fits the data slightly better than the three-class alternative, but not as well as the original four-class model. In addition, the significant LMR and BLRT test results indicate that the two-class model provides better fit than a one-class model. Table 7 contains the threshold and proportion of individuals endorsing each item. The response patterns seen herein do correspond, generally speaking, to the hypothesized patterns for the two-class solution.

Although we had a hypothesis regarding the likely number of classes present in the population, four in this case, we also examined other possible solutions. Indeed, when conducting an LCA, a researcher should be open to investigating other possible models in addition to

TABLE 7
Threshold Parameter Estimates and Proportion Endorsing Items for the Two-Class Confirmatory Latent Class Analysis Model, Deterministic Confirmatory Latent Class Analysis Model, and Inequality Constrained Confirmatory Latent Class Analysis Model for Future Purpose

	Two-Class Model		*Deterministic Model*		*Inequality Constrained Model*	
Item	*Class 1 (220)*	*Class 2 (154)*	*Class 1 (271)*	*Class 2 (103)*	*Class 1 (220)*	*Class 2 (154)*
Change the way people think (y4)	−5.055/.994	5.443/.004	−0.098/.524	−15.000/1.000	−0.882/.707	0.231/.442
Create something new (y5)	−0.873/.705	0.218/.446	−0.047/.512	−15.000/1.000	−4.207/.985	4.207/.015
Make money (y10)	−0.401/.599	−0.401/.599	−0.098/.524	−15.000/1.000	−0.401/.599	−0.401/.599
Have fun (y15)	−2.335/.912	−2.335/.912	−2.335/.912	−2.335/.912	−2.335/.912	−2.335/.912

the one that he or she originally proposed, or as is common in structural equation modeling, have competing models that can be compared with one another. When comparing the models, both the relative fit as measured by statistics such as the aBIC and the LMR and BLRT tests must be considered, as well as the nature of the latent classes revealed by the analysis and their correspondence to substantive theories about how participants group together. This latter concern is similar to the way that groupings of observed variables into factors must make substantive sense for an exploratory factor analysis solution to be viable. The researcher can make such determinations by examining the variable response patterns for members of the latent classes and comparing them with what theory would predict. For the solution to have meaning, the pattern of responses on these items for each class must be theoretically viable.

DETERMINISTIC CONSTRAINTS

In addition to constraining parameter estimates in two or more groups to be equal, it is also possible to constrain thresholds to be a specific value, corresponding to McCutcheon's (2002) deterministic constraint CLCA. We can do this in M*plus* by replacing the * in the MODEL commands with @ for specific items. Whereas the * provides M*plus* with starting values for threshold estimation, @ sets the threshold value to the number immediately following it. Thus, for example, if we expect all members of one latent class to endorse a specific item, we can set the threshold so as to ensure the probability of endorsement to be 1 for this class. In the two-class model described earlier, setting the threshold of Class 2 to −15 for Item 10 ensures that all members of that class will indicate that they want to "make money." Referring to Equation 4, we can see that setting the threshold to −15 results in an endorsement probability value of essentially 1:

$$P = \frac{1}{1+e^{-\tau}} = \frac{1}{1+e^{-15}} = \frac{1}{1+0.0000003} = 0.9999997$$

The MODEL statement for estimating this model appears in the Appendix.

Using Equation 4 it is possible to translate any probability value into a threshold that could then be used in the M*plus* MODEL statement. Although in this example only one parameter value was set, it is possible to set multiple parameters to specific values to model very specific latent class structures. Such constraints allow for the assessment of very specific hypotheses regarding the latent class structure in the data. On the other hand, it should be noted that estimation of model parameters in LCA is not done independently, so that setting one or more thresholds to specific values will impact the estimation of other item threshold values as well as latent class membership. As a result of such constraints, the estimation of other model parameters might be unrealistic or apply to a very small number of participants. Therefore, much care needs be taken prior to setting these values to ensure that there is a strong theoretical basis for doing so.

The fit statistics for this deterministic model appear in Table 4. Based on the aBIC this model fits the data less well than any of the equality constraint models previously described. The significant LMR and BLRT tests reveal that this deterministic two-class model fits the data better than a one-class model. The threshold and probability of endorsement values appear in

Table 6. Neither group displays item parameter values corresponding to those that had been hypothesized, although the threshold for y10 is 1 for latent Class 2, as it was constrained to be. It is important to note, however, that by setting one parameter to an extreme value we have fundamentally changed the basic makeup of the resulting two-class solution as compared with the results for the original two-class model.

Another issue of some import is that the deterministic model is nested in the two-class model, because the former is essentially the same as the latter except for the constraint placed on variable y10. Because of this nested relationship, it is possible to statistically compare the fit of the models using the difference in their χ^2 values, much in the way that one can test for differences of model fit for nested structural equation models. In this case, the likelihood ratio χ^2 for the deterministic model was 41.623, with 8 *df*, and the likelihood ratio χ^2 for the standard two-class model was 31.449 with 7 *df*. The χ^2 difference is 10.174 with 1 *df* which yields a *p* value of .0014. Therefore we can conclude that the standard two-class model provides significantly better data fit than the deterministic two-class model.

INEQUALITY CONSTRAINTS

The previous examples demonstrated how M*plus* can be used to conduct a CLCA in which specific group parameter values are constrained to be equal, and when a latent class is hypothesized to have a specific probability value for an item. It is also possible to model specific inequalities between latent classes for thresholds of one or more variables. For example, rather than simply constraining two or more groups to have the same or different threshold values, we can be more specific in modeling one latent class to have a higher threshold value (lower probability of item endorsement) than another. For example, in the two-class model we might hypothesize that the threshold for the item "create something new" will be lower for one latent class than for the other. Using the MODEL CONSTRAINT command the researcher restricts the threshold for one class to be the negative of the threshold for the other, ensuring that one latent class will have a higher probability of endorsing the item than the other. It is also possible to constrain the parameter of one class to be higher than that of another, without restricting one to be the negative of the other. The M*plus* MODEL statement for the first analysis appears in the Appendix.

The threshold value for y5 is named **p1** in Class 1 and **p2** in Class 2. The MODEL CONSTRAINT command then establishes an explicitly directional hypothesis for the threshold values and thereby the probabilities of item endorsement. In this case, the threshold for item y5 for latent Class 1 was set equal to the negative of the threshold for latent Class 2. This constraint means that latent Class 1 will have a lower threshold and higher probability of endorsing the item than will Class 2. The fit statistics appear in Table 4 in the Inequality 1 row. This model does not fit the data as well as the four latent class solution, although it does appear to be somewhat better than the other alternatives discussed previously, based on the value of aBIC. M*plus* does not provide the LMR test when using the MODEL CONSTRAINT command. However, the BLRT is still available, with the significant value indicating that this two-class model fits the data better than a one-class model would. The threshold and probability of item endorsement values are displayed in Table 7. We constrained the threshold of the item "Create something new" for latent Class 1 to be the negative of that for latent Class 2 in the MODEL

CONSTRAINT command, and indeed the resulting output yields a threshold of −4.207 for latent Class 1 and 4.207 for latent Class 2. The corresponding proportion of individuals in Class 1 endorsing this item was 0.985, whereas for Class 2 the proportion endorsing was 0.015. The other parameter estimates suggest that latent Class 1 corresponds to those individuals who are likely to endorse all of the items on the scale, whereas latent Class 2 corresponds to those who are most likely to endorse "make money" and "have fun" as their primary purposes in life. This two-class inequality constrained model is nested within the more general two-class model so that we can compare their relative fit to the data using the χ^2 difference test described earlier. As noted previously, the χ^2 for the standard two-class model was 31.449 with 7 *df*, whereas the χ^2 for this inequality constrained model was 37.073 with 9 *df* (there were two additional constraints in this model, one involving the difference on variable y5 and the other involving the equality of variable y10). The χ^2 difference was 5.624, with 2 *df* and a p value of .06. Therefore, we would conclude that the fit of the two models was not significantly different.

Using the MODEL CONSTRAINT command, it is also possible to establish a somewhat more sophisticated ordering of threshold values for two or more groups. For example, in the three-group model, the researcher might have reason to believe that the likelihood of endorsing "have fun" differs such that the groups are ordered sequentially, with one group having the highest likelihood of item endorsement (lowest threshold), followed by the second group and then the third. Using MODEL CONSTRAINT it is possible to express this ordering using the M*plus* commands in the Appendix.

In this case, the parameter restrictions are such that the third group has a threshold twice the size of that for Class 1 on the item "have fun." In turn, Class 2 has a threshold 1.5 times that for Class 1. Other parameter restrictions could certainly be used here, but some combinations of these restrictions might not be found in the data, thus resulting in empty groups. For example, using M*plus* we could introduce constraints that a single group has threshold values that are three times as large as those of another group for the items "Make money," "Change the way people think" and "Create something new." However, in the sample as a whole it might be that no combination of individuals produced item responses that would satisfy this type of constraint, resulting in a latent class containing no individuals. The fit statistics for the model in which thresholds in one group are three times larger than those in another appear in Table 4 in the Inequality 2 row. The fit of this model is better than that of the others, which is likely because only one set of parameter constraints was imposed, as opposed to the larger number of constraints in the other models. The thresholds (proportion of endorsement) for y15 for the three classes were −3.394 (0.968), −2.546 (0.927), and −1.697 (0.845), respectively.

If a researcher believes there to be an ordered pattern of group parameters on the item "have fun," but does not wish to place the specific restrictions on the degree of difference in these values, he or she could use the following under MODEL CONSTRAINT:

```
p3>p2;
p2>p1;
```

This code requires the third class to have a higher threshold than the second class, which in turn will have a higher threshold than the first class. Unlike the previous set of commands, no constraints are placed on the magnitude of the difference between the parameters. The resulting analysis (model fit values appear in Table 4 as Inequality 2b) produced an aBIC just slightly larger than that for the previous set of constraints, and the BLRT was statistically significant, indicating that three classes indeed fit the data better than two. The threshold values (proportions of endorsement) for the three latent classes were −4.052 (0.983), −3.839 (0.979), and −0.678 (0.663).

CONCLUSIONS

CLCA is a powerful tool for testing theories regarding the nature of specific latent classes in a population. Unlike exploratory LCA, which does not incorporate a priori substantive hypotheses about latent groups in the population, CLCA allows the researcher to specify response patterns in the observed variables that correspond to what would be expected by underlying groups given a specific hypothesis. These specifications take the form of restrictions of conditional probabilities for item endorsement for different latent classes. The software package M*plus* can be used to model and assess CLCA solutions for dichotomous variables using restrictions on thresholds, which correspond directly to probabilities of a given response. The three primary types of CLCA modeling described in McCutcheon (2002), including equality constrained, deterministic, and inequality constrained models can all be analyzed using M*plus*.

Researchers employing CLCA can elect to use one or more of the modeling strategies demonstrated. The guiding factor in selecting which of these to use should be the hypothesis, based in substantive theory, that the researcher brings to the problem. For example, if the primary research hypothesis simply states that two latent classes will have an equal likelihood of endorsing an item, whereas a third class will potentially have a different such likelihood, then the equality constraints described earlier might be sufficient. On the other hand, if theory holds that one latent class will have a higher likelihood (or even more specifically be twice as likely) to endorse an item than another class, then the inequality constraints might be appropriate. If a researcher were to start with a specific hypothesis (e.g., one class is twice as likely to endorse an item as the other class) but not find empirical support for it (i.e., model fit is poor) he or she can then alter the model to be more general to ascertain if some other pattern is more likely to be present in the population. In short, a researcher might examine multiple, related hypotheses based on substantive considerations as well as empirical evidence. However, a major caveat must be made here regarding the interpretation of such exploratory analyses. If a researcher starts with specific hypotheses and then revises them to be more general given the evidence provided by CLCA, they must be extremely careful not to make definitive conclusions regarding the state of the population because the results they are seeing might well be the result of sampling variation. Any conclusions drawn from these exploratory analyses with CLCA would need to be kept tentative and used to design future studies in the area.

Given the relative ease with which these models can be estimated, it is important that researchers carefully consider the hypotheses that they want to assess. Constraining some parameters in the ways described here will have a direct impact on the estimation of other parameters. Thus, it is important that the restrictions used have a theoretical basis and that they be reasonable given the sample. Otherwise, the researcher might find that some combinations of hypotheses result in untenable results, empty latent classes, or both.

One issue that must be considered by researchers interested in using LCA in general, including CLCA models, is sample size. Early work in the area of sample size and LCA focused on the necessary sample for the chi-square test of model fit to be accurate (e.g., Fienberg, 1979; Rudas, 1986). However, subsequent research (McCutcheon, 2002) has demonstrated the limitations of using this test so that this work is no longer relevant to researchers using LCA. More recent work has focused on the model fit statistics used in this study, including the adjusted LMR, BLRT, and the aBIC. Lo, Mendell, and Rubin (2001) found that for samples of less than 300 the adjusted LMR displayed low power for detecting the correct model. Henson, Reise, and Kim (2007) reported that even with samples of 500, model fit statistics might not exhibit sufficient power for correctly detecting the presence of a two-class latent model versus one class in the conditions that they simulated. In addition to problems with accurately identifying the correct model, Henson et al. also found that using LCA with samples of 500 was associated with problems obtaining convergence when estimating parameters. Although this was not a problem with the examples demonstrated earlier, the Henson et al. simulation study included more observed variables (9) than were used here. Work by Nylund et al. (2007) produced similar results with respect to sample size. Specifically, they found that for the smallest sample size condition (200) the ability of the BLRT and aBIC statistics to correctly identify the number of latent classes was somewhat compromised, although with $n = 500$ both methods were typically very accurate. It should be noted that the underlying models used in the Henson et al. and Nylund et al. studies were somewhat different, as were the number of indicator variables. Taken together, it would appear that LCA requires samples well into the hundreds, with most simulation studies suggesting 500 as a worthy goal in practice. It should be noted again, however, that the examples reported here were based on a sample of 374, albeit with a small number of indicators.

REFERENCES

Benson, P. L. (2006). *All kids are our kids: What communities must do to raise caring and responsible children and adolescents* (2nd ed.). San Francisco, CA: Jossey Bass.

Bundick, M., Andrews, M., Jones, A., Mariano, J. M., Bronk, K. C., & Damon, W. (2006). Revised youth purpose survey. Unpublished instrument, Stanford Center on Adolescence, Stanford, CA.

Croon, M. A. (1990). Latent class analysis with ordered latent classes. *British Journal of Mathematical and Statistical Psychology, 43,* 171–192.

Damon, W. (2009). *The path to purpose: How young people find their calling in life.* New York: Free Press.

Damon, W., Menon, J., & Bronk, K. C. (2003). The development of purpose during adolescence. *Applied Developmental Science, 7*(3), 119–128.

De Vogler, K. L., & Ebersole, P. (1980). Categorization of college students' meaning in life. *Psychological Reports, 46,* 387–390.

De Vogler, K. L., & Ebersole, P. (1981). Adults' meaning in life. *Psychological Reports, 49,* 87–90.

De Vogler, K. L., & Ebersole, P. (1983). Young adolescents' meaning in life. *Psychological Reports, 52,* 427–431.

Fienberg, S. E. (1979). The use of chi-squared statistics for categorical data problems. *Journal of the Royal Statistical Society, Series B, 41,* 425–439.

Goodman, L. A. (2002). Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 3–55). Cambridge, UK: Cambridge University Press.

Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling, 14,* 202–226.

Hoijtink, H. (2001). Confirmatory latent class analysis: Model selection using Bayes factors and (pseudo) likelihood ratio statistics. *Multivariate Behavioral Statistics, 36,* 563–588.

Laudy, O., Boom, J., & Hoijtink, H. (2005). Bayesian computational methods for inequality constrained latent class analysis. In L. A. van der Ark, M. A. Croon, & K. Sijtsma (Eds.), *New developments in categorical data analysis for the social and behavioral sciences* (pp. 63–82). Mahwah, NJ: Lawrence Erlbaum Associates.

Lo, Y., Mendell, N., & Rubin, D. (2001). Testing the number of components in a normal mixture. *Biometrika, 88,* 767–778.

McCutcheon, A. L. (2002). Basic concepts and procedures in single-and multiple-group latent class analysis. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 57–88). Cambridge, UK: Cambridge University Press.

McLachlan, G., & Peel, D. (2000). *Finite mixture models.* New York: Wiley.

Muthén, B. O. (2001). Latent variable mixture modeling. In G. A. Marcoulides & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 1–34). Mahwah, NJ: Lawrence Erlbaum Associates.

Muthén, L. K., & Muthén, B. O. (2008). *M*plus *user's guide* (5th ed.). Los Angeles: Muthén & Muthén.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling, 14,* 535–569.

Rudas, T. (1986). A Monte Carlo comparison of the small sample behavior of the Pearson, the likelihood ratio and the Cressie-Read statistics. *Journal of Statistical Computation and Simulation, 24,* 107–120.

Showalter, S. M., & Wagener, L. M. (2000). Adolescents' meaning in life: A replication of De Vogler and Ebersole (1983). *Psychological Reports, 87,* 115–126.

APPENDIX

M*plus* Commands for LCA with Equality Constraints

```
TITLE:    CLCA for 4 classes of adolescent purpose
DATA:     FILE IS all.dat;
VARIABLE: NAMES ARE id y1-y17;
     USEVARIABLES ARE y4 y5 y10 y15;
  CATEGORICAL ARE y4 y5 y10 y15;
     CLASSES = c (4);
ANALYSIS: TYPE = MIXTURE;
    STARTS=100 10;
    LRTBOOTSTRAP=100;
 MODEL: %overall%
    %c#1%
    [y4$1*-2] (1);
    [y5$1*-2] (3);
    [y10$1*2] (5);
    [y15$1*-2] (7);
    %c#2%
    [y4$1*2] (2);
    [y5$1*2] (4);
    [y10$1*2] (5);
    [y15$1*-2] (7);
    %c#3%
    [y4$1*2] (2);
    [y5$1*2] (4);
    [y10$1*-2] (6);
    [y15$1*-2] (7);
    %c#4%
    [y4$1*-2] (1);
    [y5$1*-2] (3);
```

```
    [y10$1*-2] (6);
    [y15$1*-2] (7);
OUTPUT: TECH11 TECH14;
```

M*plus* Code for Estimating Alternate Equality Constraint Model with Three Classes Rather Than Four

```
MODEL: %overall%
   %c#1%
   [y4$1*-2] (1);
   [y5$1*-2] (3);
   [y10$1*2] (5);
   [y15$1*-2] (7);
   %c#2%
   [y4$1*-2] (1);
   [y5$1*-2] (3);
   [y10$1*-2] (6);
   [y15$1*-2] (7);
   %c#3%
   [y4$1*2] (2);
   [y5$1*2] (4);
   [y10$1*2] (6);
   [y15$1*-2] (7);
```

M*plus* Code for Constraining One Group to Have All Members Endorse Item y10

```
MODEL: %overall%
   %c#1%
   [y4$1*-2];
   [y5$1*-2];
   [y10$1*-2];
   [y15$1*-2] (6);
   %c#2%
   [y4$1*2];
   [y5$1*2];
   [y10$1@-15];
   [y15$1*-2] (6);
```

M*plus* Code for Constraining the Threshold of Class 2 to Be the Negative of the Threshold for Class 1

```
MODEL: %overall%
   %c#1%
   [y4$1*-2] (3);
   [y5$1*-2] (p1);
   [y10$1*-2] (5);
   [y15$1*-2] (6);
   %c#2%
```

```
   [y4$1*2] (4);
   [y5$1*2] (p2);
   [y10$1*-2] (5);
   [y15$1*-2] (6);
MODEL CONSTRAINT:
   p1=-p2;
```

M*plus* Code for Ordering the Thresholds of Three Latent Classes

```
MODEL: %overall%
   %c#1%
   [y4$1*-2] ;
   [y5$1*-2] ;
   [y10$1*2] ;
   [y15$1*-2] (p1);
   %c#2%
   [y4$1*-2] ;
   [y5$1*-2] ;
   [y10$1*-2] ;
   [y15$1*-2] (p2);
   %c#3%
   [y4$1*2] ;
   [y5$1*2] ;
   [y10$1*-2];
   [y15$1*-2] (p3);
MODEL CONSTRAINT:
   p3=2*p1;
   p2=1.5*p1;
```