# 6 Supplementary Material

## 6.1 Optimization procedure of ICQF

Recall that the Lagrangian $\mathcal{L}_\rho$ of ICQF is:

$$\mathcal{L}_\rho(W,Q,Z,\alpha_Z) = \frac{1}{2}\|\mathcal{M}\odot(M-Z)\|_F^2 + \mathcal{I}_{\mathcal{W}}(W) + \beta\|W\|_{1,1} + \mathcal{I}_{\mathcal{Q}}() + \beta\|Q\|_{1,1} \tag{13}$$

$$+ \left\langle \alpha_Z, Z - [W,C]Q^T \right\rangle + \frac{\rho}{2}\left\|Z - [W,C]Q^T\right\|_F^2 + \mathcal{I}_{\mathcal{Z}}(Z) \tag{14}$$

Following the ADMM approach, we alternatingly update primal variables $W, Q$ and the auxiliary variable $Z$, instead of updating them jointly. In particular, we iteratively solve the following sub-problems:

$$W^{(i+1)} = \operatorname*{arg\,min}_{W\in\mathcal{W}} \frac{\rho}{2}\left\|Z^{(i)} - [W,C]Q^{(i),T} + \frac{1}{\rho}\alpha_Z^{(i)}\right\|_F^2 + \beta\|W\|_{1,1} \tag{Sub-problem 1}$$

$$Q^{(i+1)} = \operatorname*{arg\,min}_{Q\in\mathcal{Q}} \frac{\rho}{2}\left\|Z^{(i)} - [W^{(i+1)},C]Q^{T} + \frac{1}{\rho}\alpha_Z^{(i)}\right\|_F^2 + \beta\|Q\|_{1,1} \tag{Sub-problem 2}$$

$$Z^{(i+1)} = \operatorname*{arg\,min}_{Z\in\mathcal{Z}} \frac{1}{2}\|\mathcal{M}\odot(M-Z)\|_F^2 + \frac{\rho}{2}\left\|Z - [W^{(i+1)},C]Q^{(i+1),T} + \frac{1}{\rho}\alpha_Z^{(i)}\right\|_F^2 \tag{Sub-problem 3}$$

for some penalty parameter $\rho$. We denote the Hadamard product as $\odot$. The vector of Lagrangian multipliers $\alpha_Z$ is updated via

$$\alpha_Z^{(i+1)} \leftarrow \alpha_Z^{(i)} + \rho(Z^{(i+1)} - [W^{(i+1)},C](Q^{(i+1)})^T) \tag{15}$$

**Sub-problems 1 and 2 (Equations 2 and 3)**

Note that equation 2 (and similarly equation 3 by taking the transpose) can be split into row-wise constrained Lasso problem. Specifically, the $r^{\text{th}}$ row problem can be simplified into:

$$x^* = \operatorname*{arg\,min}_{0\le x_i\le 1} \frac{\rho}{2}\|b-Ax\|_F^2 + \beta\|x\|_1, \quad A = Q^{(i)}, \quad b = \left[Z^{(i)} - CQ^{(i),T} + \frac{1}{\rho}\alpha_Z^{(i)}\right]_{[r,:]} \tag{16}$$

Here we use the Matlab matrix notation $\left[\cdot\right]_{[r,:]}$ to represent row extraction operation. As suggested in Gaines et al. (2018) one can also use ADMM to solve equation 16:

$$x^{(i+1)} = \arg\min \frac{\rho}{2}\|b-Ax\|_2^2 + \frac{\tau}{2}\|x - y^{(i)} + \frac{1}{\tau}\mu^{(i)}\|_2^2 + \beta\|x\|_1 \tag{17}$$

$$y^{(i+1)} = Proj_{[0,1]}(x^{(i+1)} + \frac{1}{\tau}\mu^{(i)}) \tag{18}$$

$$\mu^{(i+1)} \leftarrow \mu^{(i)} + \tau(x^{(i+1)} - y^{(i+1)}) \tag{19}$$

Similarly, $\mu$ is the vector of Lagrangian multipliers and $\tau$ is the penalty parameter. $Proj_{[0,1]}$ refers to the orthogonal projection into $[0,1]$ (inherited from the box-constraints of $W$). Equation 17 can be solved via the well-established FISTA algorithm (Beck & Teboulle, 2009). Consider the following optimization problem

$$\operatorname*{arg\,min}_x \ \lambda\|x\|_1 + \frac{1}{2}f(x) \tag{20}$$

The FISTA algorithm for solving 20 is summarized as follows:

To solve equation 17 with FISTA algorithm, using the notation as introduced in equation 16, we have

$$f(x) = \rho\|b-Ax\|_2^2 + \tau\|x - y^{(i)} + \frac{1}{\tau}\mu^{(i)}\|_2^2 \tag{21}$$

**Algorithm 1:** FISTA for equation 20

**Initialize:** $\delta = 1\text{e}{-6}$; $x_{-1} = \mathbf{0}, x_0 = t_0 = \mathbf{1}$
**Input:** $L$, Lipschitz constant of $\nabla f$
**Result:** Solution $x$ of equation 20
**while** $\|x_i - x_{i-1}\|_2 > \delta$ **do**
  $\widetilde{x}_{i+1} =_z \left\{ \frac{\lambda}{L}\|z\|_1 + \frac{1}{2}\left\|z - \left(x_i - \frac{1}{L}\nabla f(x_i)\right)\right\| \right\}$;
  $t_{i+1} = \frac{\mathbf{1}+\sqrt{\mathbf{1}+4t_i^2}}{2}$;
  $x_{i+1} = \widetilde{x}_{i+1} + \frac{t_i - \mathbf{1}}{t_{i+1}}(\widetilde{x}_{i+1} - x_i)$;
**end**

To compute $L$, the Lipschitz constant of $\nabla f$, we have

$$\begin{aligned} \nabla f(x) &= 2\rho\left(A^T A(x-b) + \tau(x-c)\right) \\ &= 2(\rho A^T A + \tau I)x - 2(\rho A^T Ab + \tau c) \end{aligned} \tag{22}$$

where $c = y^{(i)} - \frac{1}{\tau}\mu^{(i)}$. Thus, $L$ is just equal to the largest eigenvalue of $2(\rho A^T A + \tau I)$.

As recommended in Huang et al. (2016), ADMM provides flexibility to use various types of loss functions and regularizations without changing the procedure. For example, we can simply change to $L_{2,1}$ norm and equation 16 becomes a constrained ridge-regression problem, which can be efficiently solved by non-negative quadratic programming algorithms. For most clinical usage, the size of questionnaire data is manageable on a single machine. However, if optimal computational and memory efficiency is required, various stochastic optimization approaches such as Mairal et al. (2010) can replace the ADMM procedure. Yet, an unbiased sampling scheme for generating random batches that handles missing responses is also needed. Such a scheme is non-trivial to obtain, especially under the multi-questionnaires scenario.

### Sub-problem 3 (Equation 4)

Since both terms in equation 4 are in Frobenius-norm, $Z$ can be optimized entry-wise. In particular, we have the following closed-form solution for $Z^{(i+1)}$:

$$Z^{(i+1)} = \underset{[\min(M),\max(M)]}{Proj}\left(\mathcal{M} \odot M + \rho[W^{(i+1)}, C](Q^{(i+1)})^T - \alpha_Z^{(i)}\right) \oslash (\rho\mathbb{1} + \mathcal{M}) \tag{23}$$

where $\mathbb{1}$ is a 1-matrix with appropriate dimension and $\oslash$ is the Hadamard division.

## 6.2 Details and proof of Preposition 3.1

In the following, we provide a self-contained convergence proof and show that, under an appropriate choice of the penalty parameter $\rho$, the ADMM optimization scheme discussed in Section 3.2 converges to a local minimum. To simplify notation, we denote $\mathbb{V}^{(i,j,k)} = \{W^{(i)}, Q^{(j)}, Z^{(k)}\}$ to be the tuple of variables $W, Q$ and $Z$ during iteration $(i), (j)$ and $(k)$ respectively. If $i = j = k$, we abbreviate it as $\mathbb{V}^{(i)}$. We also denote $R^{(i)} = [W^{(i)}, C](Q^{(i)})^T$ and for any matrices $A, B$ with appropriate dimensions, $\langle A, B\rangle = \text{Trace}(A^T B)$. In the following, we are going to show that the Lagrangian is decreasing across iterations. Particularly, we consider the difference of Lagrangian between consecutive iterations:

$$\begin{aligned} &\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i+1)}) - \mathcal{L}_\rho(\mathbb{V}^{(i)}, \alpha_Z^{(i)}) \\ =&\underbrace{\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i+1)}) - \mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i)})}_{(I)} + \underbrace{\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i)}) - \mathcal{L}_\rho(\mathbb{V}^{(i)}, \alpha_Z^{(i)})}_{(II)} \end{aligned} \tag{24}$$

Expanding term $(I)$, we have

$$\begin{aligned} \mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i+1)}) - \mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i)}) &= \left\langle \alpha_Z^{(i+1)} - \alpha_Z^{(i)}, Z^{(i+1)} - R^{(i+1)} \right\rangle \\ &= \frac{1}{\rho}\|\alpha_Z^{(i+1)} - \alpha_Z^{(i)}\|_F^2 \end{aligned} \tag{25}$$

Expanding term $(II)$, we have

$$
\begin{aligned}
&\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i)}) - \mathcal{L}_\rho(\mathbb{V}^{(i)}, \alpha_Z^{(i)}) \\
=&\overbrace{\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i)}) - \mathcal{L}_\rho(\mathbb{V}^{(i+1,i+1,i)}, \alpha_Z^{(i)})}^{(\mathcal{A})} + \overbrace{\mathcal{L}_\rho(\mathbb{V}^{(i+1,i+1,i)}, \alpha_Z^{(i)}) - \mathcal{L}_\rho(\mathbb{V}^{(i+1,i,i)}, \alpha_Z^{(i)})}^{(\mathcal{B})} \\
&+ \underbrace{\mathcal{L}_\rho(\mathbb{V}^{(i+1,i,i)}, \alpha_Z^{(i)}) - L(S^{(k)}, \alpha_Z^{(i)})}_{(\mathcal{C})}
\end{aligned} \tag{26}
$$

Expanding $(\mathcal{A})$ by the definition, we have

$$
\begin{aligned}
&\frac{1}{2}\|\mathcal{M} \odot (M - Z^{(i+1)})\|_F^2 - \frac{1}{2}\|\mathcal{M} \odot (M - Z^{(i)})\|_F^2 + \left\langle \alpha_Z^{(i)}, Z^{(i+1)} - R^{(i+1)} \right\rangle \\
&- \left\langle \alpha_Z^{(i)}, Z^{(i)} - R^{(i+1)} \right\rangle + \frac{\rho}{2}\left\| Z^{(i+1)} - R^{(i+1)} \right\|_F^2 - \frac{\rho}{2}\left\| Z^{(i)} - R^{(i+1)} \right\|_F^2 \\
=&\left\langle \mathcal{M} \odot (Z^{(i+1)} - M), \mathcal{M} \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle - \|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 \\
&+ \langle \alpha_Z^{(i)}, Z^{(i+1)} - Z^{(i)} \rangle + \rho \left\langle Z^{(i+1)} - R^{(i+1)}, Z^{(i+1)} - Z^{(i)} \right\rangle - \rho\|Z^{(i+1)} - Z^{(i)}\|_F^2 \\
=&\left\langle \mathcal{M} \odot (Z^{(i+1)} - M) + \rho \cdot Z^{(i+1)} + \alpha_Z^{(i)} - \rho R^{(i+1)}, Z^{(i+1)} - Z^{(i)} \right\rangle \\
&- \|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 - \rho\|(Z^{(i+1)} - Z^{(i)})\|_F^2 \\
&- \left\langle \mathcal{M} \odot (Z^{(i+1)} - M), (1 - \mathcal{M}) \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle
\end{aligned}
$$

Since $Z^{(i+1)}$ is the minimizer of equation 4, we have

$$
\begin{aligned}
&\left\| \mathcal{M} \odot (M - Z^{(i+1)}) \right\|_F^2 + \rho \left\| Z^{(i+1)} - R^{(i+1)} + \frac{1}{\rho}\alpha_Z^{(i)} \right\|_F^2 \\
\leq &\left\| \mathcal{M} \odot (M - Z^{(i)}) \right\|_F^2 + \rho \left\| Z^{(i)} - R^{(i+1)} + \frac{1}{\rho}\alpha_Z^{(i)} \right\|_F^2
\end{aligned}
$$

which gives

$$
\begin{aligned}
&2\left\langle \mathcal{M} \odot (Z^{(i+1)} - M), \mathcal{M} \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle - \|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 \\
\leq &- 2\left\langle \rho \cdot Z^{(i+1)} + \alpha_Z^{(i)} - \rho R^{(i+1)}, Z^{(i+1)} - Z^{(i)} \right\rangle + \rho\|Z^{(i+1)} - Z^{(i)}\|_F^2
\end{aligned}
$$

It further implies

$$
\begin{aligned}
&\left\langle \rho \cdot Z^{(i+1)} + \alpha_Z^{(i)} - \rho R^{(i+1)}, Z^{(i+1)} - Z^{(i)} \right\rangle \\
\leq &- \left\langle \mathcal{M} \odot (Z^{(i+1)} - M), \mathcal{M} \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle + \frac{1}{2}\|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 \\
&+ \frac{\rho}{2}\|Z^{(i+1)} - Z^{(i)}\|_F^2
\end{aligned}
$$

By direct substitution, we have

$$
\begin{aligned}
(\mathcal{A}) \leq &\left\langle \mathcal{M} \odot (Z^{(i+1)} - M), Z^{(i+1)} - Z^{(i)} \right\rangle \\
&- \left\langle \mathcal{M} \odot (Z^{(i+1)} - M), \mathcal{M} \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle \\
&+ \frac{1}{2}\|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)}\|_F^2 + \frac{\rho}{2}\|Z^{(i+1)} - Z^{(i)}\|_F^2 - \|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 \\
&- \rho\|(Z^{(i+1)} - Z^{(i)})\|_F^2 - \left\langle \mathcal{M} \odot (Z^{(i+1)} - M), (1 - \mathcal{M}) \odot (Z^{(i+1)} - Z^{(i)}) \right\rangle \\
= &- \frac{1}{2}\|\mathcal{M} \odot (Z^{(i+1)} - Z^{(i)})\|_F^2 - \frac{\rho}{2}\|(Z^{(i+1)} - Z^{(i)})\|_F^2 \leq -\frac{\rho}{2}\|(Z^{(i+1)} - Z^{(i)})\|_F^2
\end{aligned} \tag{27}
$$

For the second term $(\mathcal{B})$, by definition, we have,

$$
\begin{aligned}
(\mathcal{B}) =& \frac{\rho}{2}\left\|Z^{(i)} - R^{(i+1)} + \frac{1}{\rho}\alpha_Z^{(i)}\right\|_F^2 - \frac{\rho}{2}\left\|Z^{(i)} - [W^{(i+1)}, C]Q^{(i),T} + \frac{1}{\rho}\alpha_Z^{(i)}\right\|_F^2 \\
&+ \beta\|Q^{(i+1)}\|_{1,1} - \beta\|Q^{(i)}\|_{1,1} \\
=& \rho\left\langle R^{(i+1)} - Z^{(i)} - \frac{1}{\rho}\alpha_Z^{(i)}, [W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\right\rangle \\
&- \frac{\rho}{2}\left\|[W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\right\|_F^2 + \beta(\|Q^{(i+1)}\|_{1,1} - \|Q^{(i)}\|_{1,1})
\end{aligned}
$$

We recall that $Q$ is updated via solving constrained Lasso problems for every row $Q^{(i+1)}_{[r,:]}$:

$$
y = \underset{x, 0\le x}{\arg\min}\ \beta\|x\|_1 + \frac{\rho}{2}\|b - Ax\|_2^2, \quad \text{where} \quad A = [W^{(i+1)}, C], b = \left[Z^{(i)} + \frac{1}{\rho}\alpha_Z^{(i)}\right]_{[r,:]} \tag{28}
$$

One obtains $y$ if and only if there exists $g \in \partial\|y\|_1$, the sub-differential of $\|\cdot\|_1$ such that

$$
\rho A^T(Ay - b) + \beta g = \mathbf{0}. \tag{29}
$$

As $\|\cdot\|_1$ is convex, we have

$$
\|x\|_1 \ge \|y\|_1 + \langle x - y, g\rangle \tag{30}
$$

which gives

$$
\|y\|_1 - \|x\|_1 \le \left\langle y - x, \frac{\rho}{\beta}A^T(Ay - b)\right\rangle = \left\langle A(y - x), \frac{\rho}{\beta}(Ay - b)\right\rangle \tag{31}
$$

Re-substituting $x = Q^{(i),T}_{[r,:]}$, $y = Q^{(i+1),T}_{[r,:]}$, $A = [W^{(i+1)}, C]$, $b = \left[Z^{(i)} + \frac{1}{\rho}\alpha_Z^{(i)}\right]_{[r,:]}$ and sum over $r$, we have

$$
\beta\|Q^{(i+1)}\|_{1,1} - \beta\|Q^{(i)}\|_{1,1} \le -\rho\left\langle R^{(i+1)} - Z^{(i)} - \frac{1}{\rho}\alpha_Z^{(i)}, [W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\right\rangle \tag{32}
$$

Therefore, we have

$$
(\mathcal{B}) \le -\frac{\rho}{2}\left\|[W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\right\|_F^2 \tag{33}
$$

With similar argument, we can bound $(\mathcal{C})$ by

$$
(\mathcal{C}) \le -\frac{\rho}{2}\left\|[(W^{(i+1)} - W^{(i)}), C]Q^{(i),T}\right\|_F^2 \tag{34}
$$

To get an upper bound of $\|\alpha_Z^{(i+1)} - \alpha_Z^{(i)}\|_F^2$, we have

$$
\begin{aligned}
&\|\alpha_Z^{(i+1)} - \alpha_Z^{(i)}\|_F^2 \\
\le& \|Z^{(i+1)} - Z^{(i)}\|_F^2 + \|R^{(i+1)} - R^{(i)}\|_F^2 \\
\le& \|Z^{(i+1)} - Z^{(i)}\|_F^2 + \|[W^{(i+1)}, C]Q^{(i+1),T} - [W^{(i+1)}, C]Q^{(i),T}\|_F^2 \\
&+ \|[W^{(i+1)}, C]Q^{(i),T} - [W^{(i)}, C]Q^{(i),T}\|_F^2 \\
\le& \|Z^{(i+1)} - Z^{(i)}\|_F^2 + \|[W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\|_F^2 + \|[(W^{(i+1)} - W^{(i)}), C]Q^{(i),T}\|_F^2
\end{aligned} \tag{35}
$$

Combining equation 25, 35, 26, 27, 33 and 34 with equation 24, we have

$$
\begin{aligned}
&\mathcal{L}_\rho(\mathbb{V}^{(i+1)}, \alpha_Z^{(i+1)}) - \mathcal{L}_\rho(\mathbb{V}^{(i)}, \alpha_Z^{(i)}) \\
\le& \frac{1}{\rho}\left\|\alpha_Z^{(i+1)} - \alpha_Z^{(i)}\right\|_F^2 - \frac{\rho}{2}\left\|Z^{(i+1)} - Z^{(i)}\right\|_F^2 - \frac{\rho}{2}\left\|[W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\right\|_F^2 \\
&- \frac{\rho}{2}\left\|[(W^{(i+1)} - W^{(i)}), C]Q^{(i),T}\right\|_F^2 \\
\le& \left(\frac{1}{\rho} - \frac{\rho}{2}\right)\cdot\Big(\|Z^{(i+1)} - Z^{(i)}\|_F^2 + \|[W^{(i+1)}, C](Q^{(i+1),T} - Q^{(i),T})\|_F^2 \\
&\qquad + \|[(W^{(i+1)} - W^{(i)}), C]Q^{(i),T}\|_F^2\Big).
\end{aligned} \tag{36}
$$

We set $\rho = 3$ in all experiments for sufficiency.

## 6.3 Details and proof of Preposition 3.2

Assume that there is a ground-truth factorization $(\mathbf{W}^*, \mathbf{Q}^*)$ of the given $\mathbf{M} = \mathbf{W}^*(\mathbf{Q}^*)^T$, with latent dimension $k^*$, where $\mathbf{W}^*$ and $\mathbf{Q}^*$ are matrix-valued random variables with entries sampled from some bounded distributions. With high probability, the error $\|\mathbf{M} - \mathbf{W}\mathbf{Q}^T\|_F^2$ we are minimizing is star-convex towards $(\mathbf{W}^*, \mathbf{Q}^*)$ whenever $k = k^*$ (Bjorck et al., 2021). To demonstrate the importance of the choice of $k$, we consider the scenario when $k \neq k^*$ below.

First, a more precise assumption for ICQF is to model $\mathbf{W}$ as *row-independent bounded random matrices*. Recall that $W$ is generated by arranging $n$ participants' latent representation as rows of $n \times k$ matrix, where the $n$ participants are assumed to be independent from each other and their corresponding latent representations follow a high-dimensional bounded distribution.

Second, let $(\mathbf{W}_1, \mathbf{Q}_1)$ and $(\mathbf{W}_2, \mathbf{Q}_2)$ be two factorizations with dimensions $k_1$ and $k_2$ respectively. Consider that there exists two factorizations which achieve the same critical point, i.e. **(a)**: equivalent mismatching loss in expectation, and **(b)**: equivalent expectation approximation to data matrix $\mathbf{M}$:

$$\textbf{(a)}: \ \mathbb{E}\left[\|\mathbf{M} - \mathbf{W}_1\mathbf{Q}_1^T\|_F^2\right] = \mathbb{E}\left[\|\mathbf{M} - \mathbf{W}_2\mathbf{Q}_2^T\|_F^2\right] \quad \text{and} \quad \textbf{(b)}: \ \mathbb{E}[\mathbf{W}_1\mathbf{Q}_1^T] = \mathbb{E}[\mathbf{W}_2\mathbf{Q}_2^T]$$

We also assume **(c)**: $\mathbb{E}\left[\sum_{j=1}^n (\mathbf{W}_i)_{j\kappa}^2\right] := \sigma_{\mathbf{W}_i}^2$ and $\mathbb{E}\left[\sum_{j=1}^m (\mathbf{Q}_i)_{j\kappa}^2\right] := \sigma_{\mathbf{Q}_i}^2$ for all $\kappa = k_i$, $i = 1, 2$.

Expanding **(a)**, we have

$$\mathbb{E}\left[\text{Trace}\left((\mathbf{M} - \mathbf{W}_1\mathbf{Q}_1^T)^T(\mathbf{M} - \mathbf{W}_1\mathbf{Q}_1^T)\right)\right] = \mathbb{E}\left[\text{Trace}\left((\mathbf{M} - \mathbf{W}_2\mathbf{Q}_2^T)^T(\mathbf{M} - \mathbf{W}_2\mathbf{Q}_2^T)\right)\right]$$

This gives

$$\mathbb{E}\left[\text{Trace}\left(\mathbf{W}_1^T\mathbf{W}_1\mathbf{Q}_1^T\mathbf{Q}_1 - 2\mathbf{M}^T\mathbf{W}_1\mathbf{Q}_1^T\right)\right] = \mathbb{E}\left[\text{Trace}\left(\mathbf{W}_2^T\mathbf{W}_2\mathbf{Q}_2^T\mathbf{Q}_2 - 2\mathbf{M}^T\mathbf{W}_2\mathbf{Q}_2^T\right)\right]$$

Denote $\mathbb{E}[\mathbf{W}_i] = \mu_{\mathbf{W}_i}, \mathbb{E}[\mathbf{Q}_i] = \mu_{\mathbf{Q}_i}$ for $i = 1, 2$, we have $\mathbf{W}_i = \bar{\mathbf{W}}_i + \mu_{\mathbf{W}_i}$ and $\mathbf{Q}_i = \bar{\mathbf{Q}}_i + \mu_{\mathbf{Q}_i}$, where $\bar{\mathbf{W}}_i$ and $\bar{\mathbf{Q}}_i$ denote the corresponding centered variables. Note that by the independence of $\mathbf{W}_i$ and $\mathbf{Q}_i$ and linearity of trace and expectation operator,

$$\begin{aligned}
&\mathbb{E}\left[\text{Trace}\left(\mathbf{M}^T\mathbf{W}_1\mathbf{Q}_1^T\right)\right] \\
=&\mathbb{E}\left[\text{Trace}\left(\mathbf{M}^T\bar{\mathbf{W}}_1\bar{\mathbf{Q}}_1^T + \mathbf{M}^T\bar{\mathbf{W}}_1\mu_{\mathbf{Q}_1}^T + \mathbf{M}^T\mu_{\mathbf{W}_1}\bar{\mathbf{Q}}_1^T + \mathbf{M}^T\mu_{\mathbf{W}_1}\mu_{\mathbf{Q}_1}^T\right)\right] \\
=&\text{Trace}(\mathbf{M}^T\mathbb{E}[\mathbf{W}_1]\mathbb{E}[\mathbf{Q}_1^T]) = \text{Trace}(\mathbf{M}^T\mathbb{E}[\mathbf{W}_2]\mathbb{E}[\mathbf{Q}_2^T]) = \mathbb{E}\left[\text{Trace}\left(\mathbf{M}^T\mathbf{W}_2\mathbf{Q}_2^T\right)\right]
\end{aligned} \tag{37}$$

which yields

$$\mathbb{E}\left[\text{Trace}\left(\mathbf{W}_1^T\mathbf{W}_1\mathbf{Q}_1^T\mathbf{Q}_1\right)\right] = \mathbb{E}\left[\text{Trace}\left(\mathbf{W}_2^T\mathbf{W}_2\mathbf{Q}_2^T\mathbf{Q}_2\right)\right] \tag{38}$$

Consider $\mathbb{E}\left[\text{Trace}\left(\mathbf{W}_1^T\mathbf{W}_1\mathbf{Q}_1^T\mathbf{Q}_1\right)\right]$ via definition, we have

$$\begin{aligned}
&\mathbb{E}\left[\text{Trace}\left(\mathbf{W}_1^T\mathbf{W}_1\mathbf{Q}_1^T\mathbf{Q}_1\right)\right] \\
=&\text{Trace}\left(\mathbb{E}\left[\mathbf{W}_1^T\mathbf{W}_1\right]\mathbb{E}\left[\mathbf{Q}_1^T\mathbf{Q}_1\right]\right) \\
=&\text{Trace}\left(\mathbb{E}\begin{bmatrix} \left(\sum_{j=1}^n (\mathbf{W}_1)_{j1}^2\right) & & * \\ & \ddots & \\ * & & \left(\sum_{j=1}^n (\mathbf{W}_1)_{jk_1}^2\right) \end{bmatrix} \right. \\
&\qquad \left. \times \ \mathbb{E}\begin{bmatrix} \left(\sum_{j=1}^m (\mathbf{Q}_1)_{j1}^2\right) & & 0 \\ & \ddots & \\ 0 & & \left(\sum_{j=1}^m (\mathbf{Q}_1)_{jk_1}^2\right) \end{bmatrix}\right) \\
=&\sum_{\kappa=1}^{k_1}\mathbb{E}\left[\sum_{j=1}^n (\mathbf{W}_1)_{j\kappa}^2\right]\mathbb{E}\left[\sum_{j=1}^m (\mathbf{Q}_1)_{j\kappa}^2\right]
\end{aligned} \tag{39}$$

Incorporating assumption **(c)**, we have

$$\mathbb{E}\left[\text{Trace}\left(\mathbf{W}_1^T\mathbf{W}_1\mathbf{Q}_1^T\mathbf{Q}_1\right)\right] = k_1\sigma_{\mathbf{W}_1}^2\sigma_{\mathbf{Q}_1}^2 \tag{40}$$

Consider equation 38 with $k_2 > k_1$. For $\mathbf{W}_1, \mathbf{Q}_1$, W.L.O.G. we pad $k_2 - k_1$ columns of zeros. Moreover, let $\mathbf{P}$ be an optimal $k_2 \times k_2$ permutation matrix, we also have

$$\mathbb{E}\left[\text{Trace}\left((\mathbf{W}_2\mathbf{P})^T\mathbf{W}_2\mathbf{P}(\mathbf{Q}_2\mathbf{P})^T\mathbf{Q}_2\mathbf{P}\right)\right] = \mathbb{E}\left[\text{Trace}\left(\mathbf{W}_2^T\mathbf{W}_2\mathbf{Q}_2^T\mathbf{Q}_2\right)\right] = k_2\sigma^2_{\mathbf{W}_2}\sigma^2_{\mathbf{Q}_2} \tag{41}$$

Combining with equation 38, it is equivalent to

$$k_1\sigma^2_{\mathbf{W}_1}\sigma^2_{\mathbf{Q}_1} = k_2\sigma^2_{\mathbf{W}_2}\sigma^2_{\mathbf{Q}_2} \tag{42}$$

which gives

$$\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] = \frac{\sigma^2_{\mathbf{Q}_2}}{\sigma^2_{\mathbf{Q}_1}}\mathbb{E}\left[\|\mathbf{W}_2\|_F^2\right] = \frac{\sigma^2_{\mathbf{Q}_2}}{\sigma^2_{\mathbf{Q}_1}}\mathbb{E}\left[\|\mathbf{W}_2\mathbf{P}\|_F^2\right] \tag{43}$$

To evaluate the impact of interpretability of latent representation under different latent dimension, we consider $\mathbb{E}\left[\|\mathbf{W}_1 - \mathbf{W}_2\mathbf{P}\|_F^2\right]$:

$$\begin{aligned}\mathbb{E}\left[\|\mathbf{W}_1 - \mathbf{W}_2\mathbf{P}\|_F^2\right] &= \mathbb{E}\left[\text{Trace}\left((\mathbf{W}_1 - \mathbf{W}_2\mathbf{P})^T(\mathbf{W}_1 - \mathbf{W}_2\mathbf{P})\right)\right] \\ &= \mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] + \frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] - 2\mathbb{E}\left[\text{Trace}(\mathbf{W}_1^T\mathbf{W}_2\mathbf{P})\right]\end{aligned} \tag{44}$$

As $\text{Trace}(\mathbf{W}_1^T\mathbf{W}_2P) \leq \|\mathbf{W}_1\|_F\|\mathbf{W}_2\mathbf{P}\|_F$, we also have

$$\begin{aligned}\mathbb{E}\left[\text{Trace}(\mathbf{W}_1^T\mathbf{W}_2\mathbf{P})\right] &\leq \mathbb{E}\left[\|\mathbf{W}_1\|_F\right]\cdot\mathbb{E}\left[\|\mathbf{W}_2\mathbf{P}\|_F\right] \\ &\leq \sqrt{\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right]\cdot\mathbb{E}\left[\|\mathbf{W}_2\|_F^2\right]} = \sqrt{\frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}}\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right]\end{aligned} \tag{45}$$

which implies

$$\mathbb{E}\left[\|\mathbf{W}_1 - \mathbf{W}_2\mathbf{P}\|_F^2\right] \geq \left(1 - 2\sqrt{\frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}} + \frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}\right)\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] = \left(1 - \sqrt{\frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}}\right)^2\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] \tag{46}$$

Since $\mathbf{W}_i$ is generated from row-wise independent bounded distribution, if we add a mild assumption that $\sigma^2_{\mathbf{W}_i} := \sigma^2_{\mathbf{W}}$ for all $i$ through re-scaling, Equation 42 implies $k_1\sigma^2_{\mathbf{Q}_1} = k_2\sigma^2_{\mathbf{Q}_2}$ and therefore

$$\mathbb{E}\left[\|\mathbf{W}_1 - \mathbf{W}_2\|_F^2\right] \geq \left(1 - 2\sqrt{\frac{k_2}{k_1}} + \frac{k_2}{k_1}\right)\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] = \left(\sqrt{\frac{k_2}{k_1}} - 1\right)^2\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] \tag{47}$$

If we substitute $k_1 = k^*$, $(\mathbf{W}_1, \mathbf{Q}_1) = (\mathbf{W}^*, \mathbf{Q}^*)$, we have

$$\mathbb{E}\left[\|\mathbf{W}^* - \mathbf{W}_2\|_F^2\right] \geq \left(\sqrt{\frac{k_2}{k^*}} - 1\right)^2\mathbb{E}\left[\|\mathbf{W}^*\|_F^2\right] \tag{48}$$

which means the relative expected difference between $\mathbf{W}^*$ and $\mathbf{W}_2$ is bounded below by $\left(\sqrt{\frac{k_2}{k^*}} - 1\right)^2$.

To prove that equation 48 holds in general, we consider the matrix concentration inequalities and show that large deviations from their means are exponentially unlikely. Benefitting from the model constraints, we can further assume that $W$ is generated from some high dimensional bounded distribution. In the following, we make use of the main theorem proposed in Meckes & Szarek (2012) on concentration of non-commutative random matrices polynomials. As $\mathbf{W}_i$ are generated from bounded distributions, $\|\mathbf{W}_i - \mathbb{E}[\mathbf{W}_i]\|_F$ is uniformly bounded. Therefore, it satisfies the convex concentration properties. The theorem achieves the following results:

$$\mathbb{P}\left\{\|\mathbf{W}\|_F^2 - \mathbb{E}\left[\|\mathbf{W}\|_F^2\right] > tkn^2\right\} \leq C_1\exp\left(-C_2\min(t^2, t^{1/2})n\right) \tag{49}$$

Recall that $\mathbb{E}\left[\|\mathbf{W}_1 - \mathbf{W}_2P\|_F^2\right] = \mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] + \frac{\sigma^2_{\mathbf{Q}_1}}{\sigma^2_{\mathbf{Q}_2}}\mathbb{E}\left[\|\mathbf{W}_1\|_F^2\right] - 2\mathbb{E}\left[\text{Trace}(\mathbf{W}_1^T\mathbf{W}_2\mathbf{P})\right]$. By padding $\mathbf{W}_1$ and $\mathbf{W}_2$ with zeros columns, we assume that $\mathbf{W}_i$ are all $n \times n$ matrices. Then the probability that the any one of the terms is deviating from their mean by a relative factor $\epsilon$ is less than $C_1\exp(-C_2\epsilon^2 n)$ for some small $\epsilon$. By the union bound, the probability that the either of them does is less than or equal to $C_3\exp(-C_4\epsilon^2 n)$.

### 6.4 Visualization of the experimental setup for diagnostic prediction evaluation

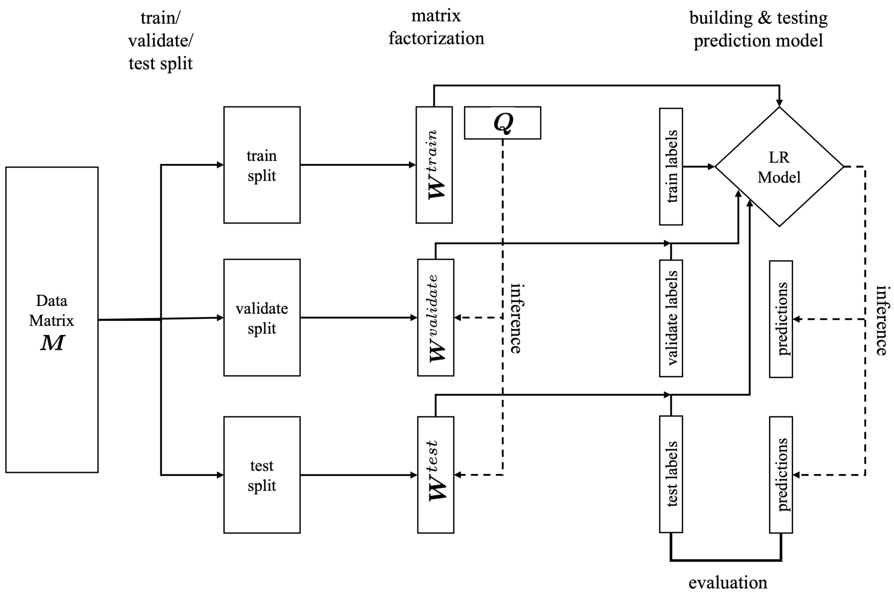


**Figure 4:** Setup for diagnostic prediction experiments.

### 6.5 Table of the 21 questionnaires used in *HBN* dataset

**Table 3:** Optimal $(k, \beta)$ of all 21 questionnaires.

| Questionnaire | Abbreviation | $n$ questions | Subscales | $k$ | $\beta$ |
|---|---|---|---|---|---|
| Affective Reactivity Index (Parent-Report) | ARI_P | 7 | nan | 2 | 0.01 |
| Affective Reactivity Index (Self-Report) | ARI_S | 7 | nan | 2 | 0.01 |
| Autism Spectrum Screening Questionnaire | ASSQ | 27 | nan | 2 | 0.01 |
| Conners 3 (Self-Report) | C3SR | 9 | | 4 | 0.05 |
| Child Behavior Checklist | CBCL | 119 | 9 | 8 | 0.5 |
| Extended Strengths and Weaknesses Assessment of Normal Behavior | ESWAN | 65 | nan | 13 | 0.2 |
| Inventory of Callous-Unemotional Traits (Parent-Report) | ICU_P | 24 | 3 | 4 | 0.1 |
| Inventory of Callous-Unemotional Traits (Self-Report) | ICU_SR | 24 | 3 | 3 | 0.1 |
| Mood and Feelings Questionnaire (Parent-Report) | MFQ_P | 34 | nan | 2 | 0.1 |
| Mood and Feelings Questionnaire (Self-Report) | MFQ_SR | 33 | nan | 2 | 0.1 |
| The Positive and Negative Affect Schedule | PANAS | 20 | 2 | 2 | 0.05 |
| Repetitive Behaviors Scale | RBS | 43 | 5 | 3 | 0.1 |
| Screen for Child Anxiety Related Disorders (Parent-Report) | SCARED_P | 41 | 5 | 3 | 0.1 |
| Screen for Child Anxiety Related Disorders (Self-Report) | SCARED_SR | 41 | 5 | 3 | 0.3 |
| Social Communication Questionnaire | SCQ | 40 | nan | 4 | 0.02 |
| Strength and Difficulties Questionnaire | SDQ | 33 | 9 | 6 | 0.05 |
| Social Responsiveness Scale (School Age) | SRS | 65 | 7 | 3 | 0.5 |
| The Strengths and Weaknesses Assessment of Normal Behavior Rating Scale for ADHD | SWAN | 18 | 2 | 3 | 0.02 |
| Symptom Checklist (Parent-Report) | SympChck | 63 | nan | 3 | 0.1 |
| Teacher Report Form (School Age) | TRF | 116 | 19 | 8 | 0.5 |
| Youth Self Report | YSR | 119 | 11 | 3 | 0.2 |

## 6.6 Full list of Top 10 questions from factorizing CBCL-HBN questionnaire

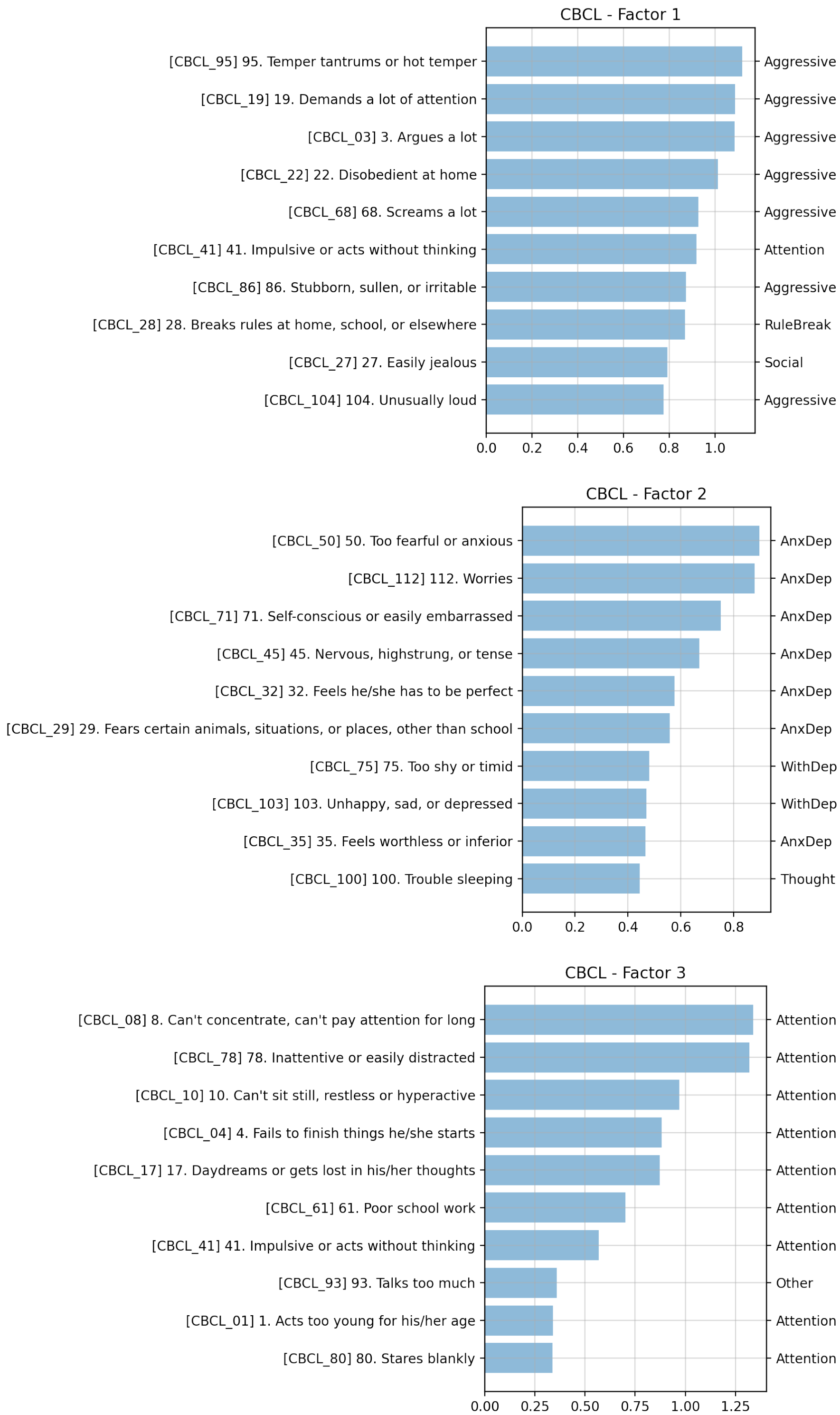


**Figure 5:** Top 10 questions ranked by $Q$ in *CBCL* using $Q$ obtained from ICQF (Factor 1-3).

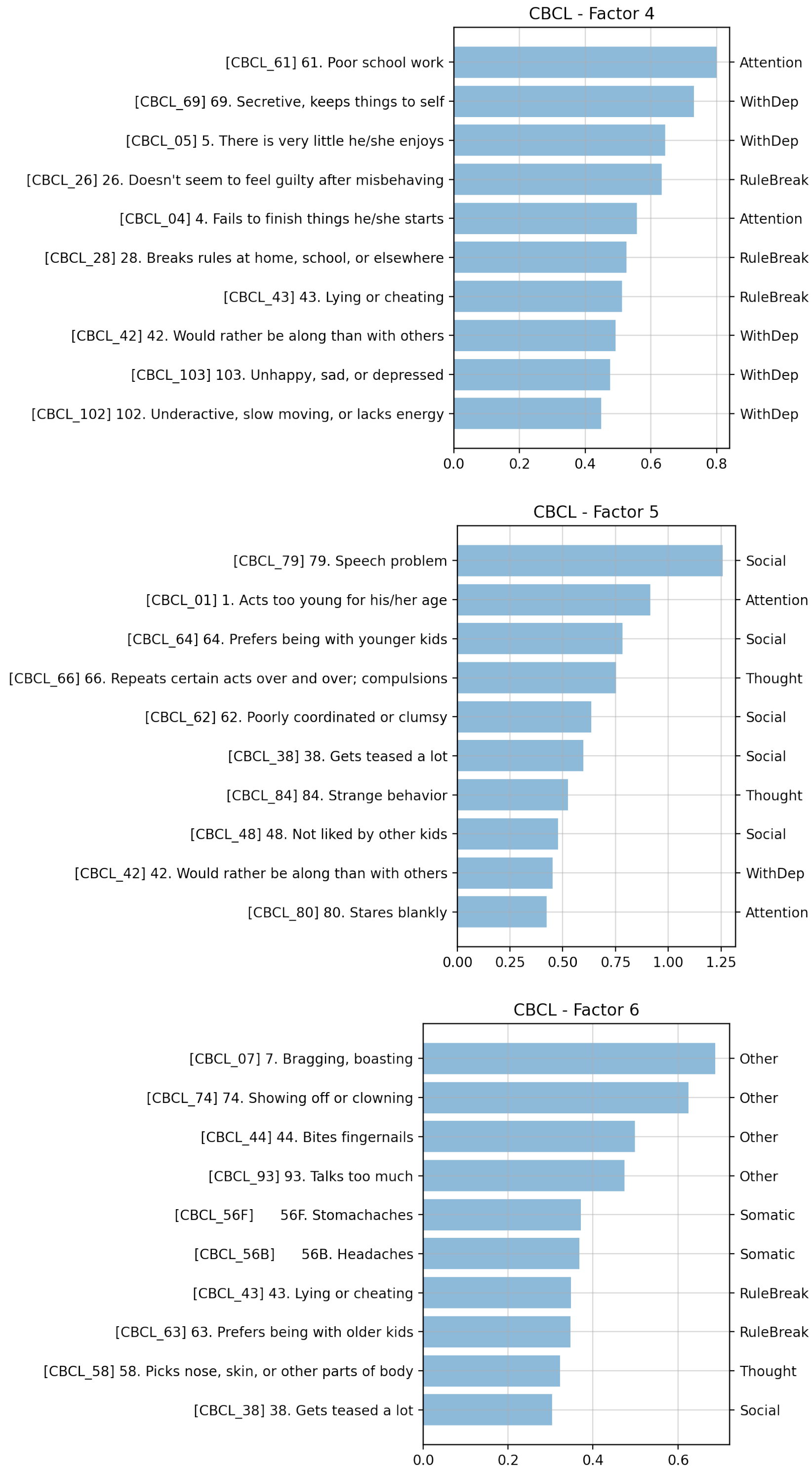


**Figure 6:** Top 10 questions ranked by $Q$ in *CBCL* using $Q$ obtained from ICQF (Factor 4-6).

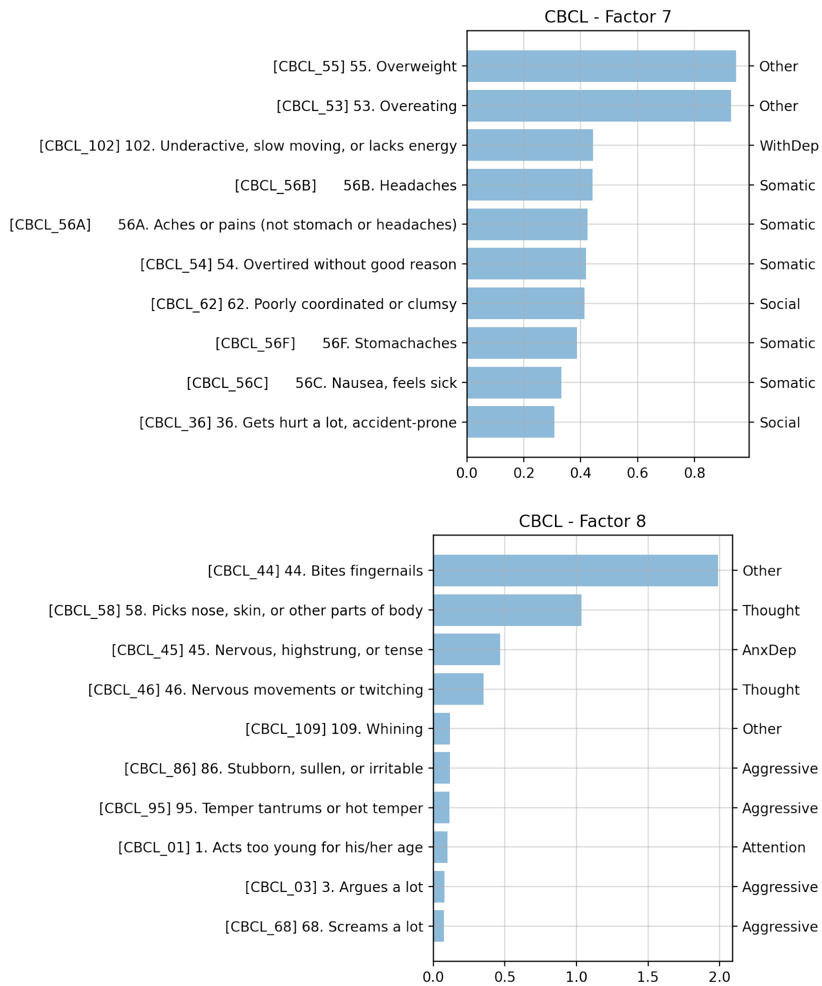


**Figure 7:** Top 10 questions ranked by $Q$ in *CBCL* using $Q$ obtained from ICQF (Factor 7-8).